

Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Física Teórica



# Filtro Óptimo para Imágenes Naturales, Aplicación al Sistema Visual

Memoria de Tesis realizada por

**D. José M. Delgado Delgado,**

presentada ante el Departamento de Física Teórica

de la Universidad Autónoma de Madrid

para optar al grado de Doctor en Ciencias (Físicas)

Trabajo dirigido por

**Dr. D. Antonio M. Turiel Martínez**

y

**Dr. D. Néstor O. Parga Carballada**

Madrid, octubre de 2007



# Índice general

<b>1. Introducción y Motivación</b>	<b>7</b>
1.1. Neurociencia computacional . . . . .	7
1.2. Procesamiento óptimo . . . . .	8
1.3. Estado del arte en el procesamiento visual . . . . .	10
1.3.1. Análisis en Componentes Independientes, Código esparso . . . . .	11
1.4. Esquema de la presentación . . . . .	12
<b>2. Filtro Óptimo para Imágenes Naturales</b>	<b>15</b>
2.1. Imágenes naturales . . . . .	15
2.2. Estadística de las imágenes naturales . . . . .	17
2.2.1. Invariancia traslacional . . . . .	18
2.2.2. Ergodicidad . . . . .	18
2.2.3. Invariancia de escala . . . . .	19
2.2.4. Espectro de potencias . . . . .	20
2.2.5. Estadística no gaussiana . . . . .	25
2.2.6. Multiescala y multifractalidad . . . . .	26
2.3. Análisis en wavelets . . . . .	28
2.3.1. Multiresolución . . . . .	30
2.4. Expansión en una base ortonormal . . . . .	30
2.5. Fórmula de la wavelet óptima . . . . .	32
2.5.1. Colectivo de imágenes utilizado . . . . .	34
2.6. Obtención de la wavelet óptima . . . . .	34

2.6.1. Validación del proceso multiplicativo . . . . .	37
2.7. Persistencia: redundancia entre escalas y componentes independientes . . . . .	39
2.8. Necesidad de más voces . . . . .	42
2.8.1. Bases orientadas . . . . .	45
2.9. Conclusiones . . . . .	49
<b>3. Taxonomía de Imágenes Naturales: Características Visuales</b>	<b>51</b>
3.1. Predominancia de patrones o características visuales . . . . .	52
3.2. Selección de colectivos y clasificación de las imágenes . . . . .	55
3.2.1. Algoritmo de clasificación de imágenes . . . . .	56
3.2.2. Extracción de las características visuales . . . . .	61
3.2.3. Filtro óptimo asociado a la característica visual . . . . .	62
3.3. Implementación del algoritmo de clasificación . . . . .	62
3.3.1. Aplicación del algoritmo en la representación de imágenes . . . . .	66
3.4. Limitación de la fórmula de la wavelet óptima: modelo analítico . . . . .	70
3.4.1. Modelo sigmoide . . . . .	73
3.4.2. Representación de imágenes utilizando el modelo . . . . .	77
3.5. Características de características visuales: Supercaracterísticas . . . . .	80
3.6. Conclusiones . . . . .	82
<b>4. Revisión de las Hipótesis para Deducir el Filtro Óptimo</b>	<b>85</b>
4.1. Estudio de los signos . . . . .	86
4.1.1. Preliminares . . . . .	86
4.1.2. Signo de $\alpha$ : $\sigma$ . . . . .	87
4.2. Medias signadas . . . . .	89
4.3. Persistencia de los signos . . . . .	92
4.3.1. Preliminares . . . . .	92
4.3.2. Signo de $\eta$ : $\beta$ . . . . .	93
4.3.3. Análisis de la persistencia . . . . .	95
4.4. Impacto de los signos en la fórmula de obtención del filtro óptimo . . . . .	96
4.5. Conclusiones . . . . .	97



<b>5. Campos Receptivos de Células Simples</b>	<b>99</b>
5.1. Del potencial de acción a los campos receptivos . . . . .	100
5.1.1. Potencial de acción . . . . .	100
5.1.2. Frecuencia de disparo . . . . .	102
5.1.3. Estimando la frecuencia de disparo . . . . .	102
5.2. Introducción al Sistema Visual Primario . . . . .	104
5.3. Campos receptivos . . . . .	107
5.3.1. Campos receptivos visuales . . . . .	107
5.3.2. Forma funcional de los campos receptivos visuales . . . . .	109
5.4. Descripción del modelo . . . . .	112
5.4.1. Simplificaciones . . . . .	113
5.4.2. Prototipos de características visuales . . . . .	114
5.4.3. Filtros . . . . .	116
5.4.4. El efecto de LGN . . . . .	117
5.5. Resultados . . . . .	118
5.5.1. Comparación con datos experimentales . . . . .	121
5.5.2. Comparación con otros modelos teóricos . . . . .	122
5.6. Conclusiones . . . . .	123
<b>6. Resumen y Conclusiones</b>	<b>125</b>
<b>A. Descomposición de Señales Infinitas, Función Unidad y Métodos de Inferencia del Filtro Óptimo</b>	<b>127</b>
A.1. Desarrollo completo de una señal infinita . . . . .	127
A.2. Aproximaciones para llegar a la wavelet óptima . . . . .	128
A.3. Métodos de inferencia . . . . .	131
<b>B. Función Gabor: Limitaciones</b>	<b>133</b>
B.1. Transformada de Fourier de la función Gabor . . . . .	133
B.2. Módulo de la transformada de Fourier . . . . .	134
B.3. Diferentes regiones . . . . .	134

B.4. Estímulos y campos receptivos . . . . .	135
<b>Índice de figuras.</b>	<b>144</b>
<b>Índice de tablas.</b>	<b>153</b>

# Capítulo 1

## Introducción y Motivación

En esta introducción vamos a presentar someramente el campo que nos ocupa, la Neurociencia computacional y, más concretamente el Procesamiento Visual y en qué ideas está basado. Se hará un recorrido por el estado de arte actual de los modelos más emblemáticos.

La idea de esta tesis se apoya en (o estudia) dos pilares fundamentales: propiedades estadísticas de las imágenes naturales -y cómo podemos construir un código óptimo a partir de éstas- y la estructura espacial de los campos receptivos de células simples -y cómo los podemos describir a partir del código óptimo construido a partir de las imágenes naturales-. La *tesis* de nuestra tesis es que existe un nexo entre los estímulos visuales y las células receptoras de éstos y a partir del estudio -estadístico- de los primeros podremos inferir propiedades de las segundas. El carácter dual del filtro (representador/analizador) es el que hace de nexo.

Acabaremos exponiendo un resumen del esquema que vamos a seguir en la presentación de este trabajo.

### 1.1. Neurociencia computacional

Dentro de la ciencia que se ocupa del sistema nervioso o de cada uno de sus diversos aspectos y funciones especializadas, esto es: la Neurociencia, existe una disciplina

que estudia su (posible) modelización matemática: la Neurociencia computacional. Su carácter, sin embargo, es multidisciplinar al tratar diversos campos de Ciencias fundamentales como la Biología, la Química o, incluso, la Física.

Desde hace tiempo es sabido [1] que las neuronas sensoriales se adaptan, a través de procesos epigenéticos<sup>1</sup>, a las propiedades estadísticas de los estímulos a los que están expuestas dichas neuronas. Attneave (1954) [2] y Barlow (1961) [1] pusieron de manifiesto que la Teoría de la Información podría establecer el vínculo matemático entre los estímulos naturales y las respuestas de las neuronas a éstos, a través del concepto de eficiencia en el código neuronal. La Teoría de la Información nos va a permitir cuantificar además las simetrías estadísticas que posea nuestra señal<sup>2</sup>, y la probabilidad de que ocurra un evento.

A lo largo de esta presentación nos vamos a centrar en los estímulos visuales para poder entender el procesamiento visual. Históricamente la percepción visual ha suscitado mucha curiosidad en general, y en particular entre investigadores de Neurociencia. En parte es debido a que los estímulos visuales tienen propiedades muy intuitivas, fáciles de asimilar por el sentido común. Es por ello que ha sido el objeto de estudio en la Neurociencia más investigado, tanto en el plano experimental como en el teórico, y sobre el que a día de hoy se tienen más resultados importantes. Muchos de éstos han sido extrapolados a otros sistemas sensoriales con relativo éxito.

## 1.2. Procesamiento óptimo

La Teoría de la Información fue desarrollada en las décadas de los cuarenta y cincuenta del siglo pasado en respuesta a la necesidad de diseñar mecanismos eficientes de comunicación. Esta teoría, a pesar de su motivación original, es una teoría matemática que profundiza en los aspectos más formales del proceso de la comunicación. Esta

---

<sup>1</sup>Epigénesis: Doctrina según la cual los rasgos que caracterizan a un ser vivo se configuran en el curso del desarrollo, sin estar preformados en el huevo fecundado. (D.R.A.E.)

<sup>2</sup>Por señal consideramos una variable aleatoria con una distribución que intenta modelizar al estímulo natural.

teoría cuantifica una señal de la siguiente manera. Consideremos el conjunto  $M$  de todas las señales posibles (las imágenes naturales, los sonidos naturales ...). Definimos la *información* asociada a una señal como

$$\mathcal{I}(w) \equiv -\log_2 \mathcal{P}(w) , \quad (1.2.1)$$

donde  $\mathcal{P}(w)$  es la probabilidad asociada a la señal  $w$ . Esta *información* se expresa en **bits**. Las señales que sean más raras, menos probables, tendrán un contenido de información mayor que otras señales que sean más frecuentes.

Si hacemos el promedio de todas estas cantidades de colectivo  $M$  tendremos

$$\mathcal{H}(M) \equiv \sum_{\{M\}} \mathcal{P}(w) \mathcal{I}(w) = \sum_{\{M\}} -\mathcal{P}(w) \log_2 \mathcal{P}(w) , \quad (1.2.2)$$

que se conoce como entropía o promedio de la información por señal.

Las señales no son independientes unas de otras y una misma señal, incluso, puede tener autocorrelaciones. Esto provoca que haya un cierto grado de redundancia. A la hora de construir un código eficiente tendremos en cuenta esta redundancia, de tal forma que el código posea la misma información que las señales originales pero minimizando esta redundancia. Esta es la idea que hay detrás del código factorial, y que fue desarrollada por Barlow [1] . Las componentes de un código factorial son variables estadísticamente independientes, que tendrán en general complicadas dependencias funcionales con las señales. El procesamiento neuronal utiliza esta estrategia para extraer las distintas componentes independientes que poseen los estímulos y así codificar de forma compacta la información que le llega de estos estímulos.

Otra posible estrategia de optimización es construir un código que maximice la cantidad de información que puede codificar dado un colectivo de señales. Esto sugiere la posible adaptación que han tenido que hacer las células sensoriales para adecuarse al entorno natural. Esta idea fue desarrollada por Linsker [3] bajo el nombre de principio *infomax* en un modelo de las primeras capas del sistema visual.

Se ha comprobado, que bajo ciertas condiciones ([4]), el principio *infomax* de Linsker y la reducción de la redundancia de Barlow, son dos formas de expresar el mismo fenómeno: optimización de la información; y es por ello que conducen a las mismas predicciones.

### 1.3. Estado del arte en el procesamiento visual

A pesar de que hoy en día es comúnmente aceptado en Neurociencia la idea de que el procesamiento neuronal tiene que estar influido por la estadística del medio, ha sido harto difícil establecer el vínculo de manera cuantitativamente precisa. Hace ya 40 años, Attneave, motivado por los trabajos que se estaban haciendo sobre Teoría de la Información, propuso que la meta que debe seguir el procesamiento visual primario<sup>3</sup> es la realización de una representación eficiente de la señal. Barlow, por su parte, hizo la hipótesis de que el papel que desempeñan las neuronas sensoriales es la de extraer la redundancia estadística de los estímulos de entrada. Variaciones sobre la *eficiencia de la codificación* se han venido sucediendo de la mano de numerosos autores ([5],[6],[7],[8] y [9]).

Pero aún con el vínculo establecido, la hipótesis no está plenamente definida. Uno necesita además establecer qué parte del entorno va a influir en el sistema. De forma cuantitativa, necesitamos especificar la distribución de probabilidad de la señal. Claro que esto es ya un problema en sí mismo, algunos autores han basado sus estudios en la observación estadística de un conjunto de imágenes, siendo éstas representativas de un entorno. Así mismo, la escala temporal también debe ser tomada en cuenta a la hora de estudiar el estímulo. Finalmente, uno necesita saber qué neuronas satisfarán el criterio de eficiencia, y cómo habrá que interpretar sus respuestas.

Básicamente existen dos corrientes metodológicas que tratan las hipótesis que antes hemos hablado. El tratamiento más directo consiste en examinar las propiedades estadísticas de las respuestas neuronales bajo estímulos naturales; este tratamiento dará los modelos experimentales (vgr. [5] y [9]). La otra corriente, donde tendrán cabida los modelos teóricos, está basada en la derivación de un modelo plausible para el procesamiento visual primario (como por ejemplo [6],[10],[11],[12] y [13]). En ella, a través del estudio de las propiedades estadísticas del medio, se deriva una representación que sea capaz de reproducir las respuestas en neuronas sensoriales habiendo

---

<sup>3</sup>Este tiene lugar en las primeras partes del sistema visual: la retina, el núcleo geniculado lateral (LGN) -localizado en el tálamo dorsal- y la corteza visual primaria (V1) -situada en la parte occipital de la corteza-.

impuesto criterios de optimización estadística.

### 1.3.1. Análisis en Componentes Independientes, Código esparso

El estudio de las componentes independientes incluye una serie de modelos que tratan de descomponer las señales de los estímulos en un conjunto de variables independientes. Un código que esté formado por variables independientes reducirá la redundancia y, por tanto, será más eficiente [14]. Para hallar estas variables independientes se aplican determinadas transformaciones a la señal. Si estas transformaciones son lineales, y por tanto sólo se basan en la estadística de segundo orden, la mejor técnica a aplicar en este caso es la denominada como Análisis de las Componentes Principales, (PCA). Esta técnica no es efectiva cuando analizamos señales que se alejan de la linealidad, por ello hoy en día no se usa para tratar imágenes naturales.

Una modificación de la técnica anterior es el Análisis de Componentes Independientes, (ICA). Esta técnica sí tiene en cuenta la estadística no gaussiana de las señales, por tanto es una buena candidata para el estudio de las imágenes naturales. El algoritmo en que se basa esta técnica, para la extracción de las componentes independientes de la señal, es la maximización de la matriz formada por los momentos de orden superior a dos, para lo que típicamente se suele elegir la matriz de curtosis. Esta técnica es muy difícil de implementar en las imágenes, por ello se suelen hacer aproximaciones. El lector interesado puede consultar las predicciones de esta técnica aplicada al campo visual en [15], en este artículo los autores hacen una aproximación lineal de ICA a las imágenes naturales, y aunque esta aproximación contradice la estadística no lineal de las imágenes, obtienen unos buenos resultados.

Por Código Esparso se entiende todas las técnicas que intentando deducir la forma de los campos receptivos visuales, (ver sección B.4) imponen que el código tenga una distribución esparsa. Una distribución de probabilidad de una variable aleatoria se dice que es esparsa cuando los valores que toma esta distribución son prácticamente cero en todos los puntos excepto en un intervalo (o intervalos) pequeño, por tanto la función de

distribución está muy concentrada en ese intervalo. Algunas de estas técnicas, [10, 16], relacionan la estructura entre los campos receptivos de células simples y el Código Esparso sin imponer ninguna forma funcional para estos campos receptivos.

## 1.4. Esquema de la presentación

Nuestro trabajo parte de la base de que si uno conoce propiedades de los estímulos podrá deducir cómo funciona el sistema. Por ello, habremos de explicar qué entendemos por estímulos visuales, las simplificaciones que hacemos a la hora de representarlos y las limitaciones que contiene. Todo esto se abordará en el capítulo 2 para después seguir con el estudio de las propiedades estadísticas de las imágenes naturales, fundamentales para derivar un filtro óptimo cuyo código maximiza la información minimizando la redundancia, extrayendo las componentes independientes a partir de las proyecciones a dos escalas consecutivas.

En el capítulo 3 aplicaremos el modelo que extrae el filtro óptimo a distintos colectivos de imágenes naturales de tamaño más pequeño. Determinaremos los colectivos mediante un algoritmo de clasificación según una jerarquización de ciertos patrones (bordes, barras, dobles barras, etc.) presentes en las imágenes naturales. El estudio taxonómico agrupará las imágenes naturales en diferentes clases según la característica visual o patrón que más predomine en éstas. A modo de ejemplo, se mostrarán dos métodos de reconstrucción de imágenes.

De los capítulos anteriores vemos que necesitamos hacer una revisión en profundidad de la derivación de la fórmula de obtención del filtro óptimo para poder dar solidez matemática a los argumentos intuitivos que se han ido planteando. Esto lo haremos en el capítulo 4. Una serie de inconsistencias teóricas serán también abordadas en profundidad.

En el capítulo 5 presentaremos un modelo que permite deducir la estructura de los campos receptivos de las células simples a partir del estudio de las imágenes naturales. Necesitaremos antes hacer una revisión del sistema visual y de los diversos modelos que intentan predecir la estructura de los campos receptivos.



El resumen de esta tesis, sus conclusiones y el trabajo futuro será discutido en profundidad en el último capítulo, 6.

En el apéndice A discutimos la validación de las hipótesis que hemos tenido que asumir al tratar con señales finitas. Por último, el apéndice B expone las limitaciones de la función Gabor como aproximante matemático de los campos receptivos.



## Capítulo 2

# Filtro Óptimo para Imágenes Naturales

En este capítulo vamos a definir de manera formal el objeto principal de estudio de esta tesis: las imágenes naturales. Analizaremos con detalle las simetrías estadísticas que se encuentran en éstas para así abordar una revisión al modelo desarrollado por Turiel et al ([13]). El resultado fundamental de este estudio es la existencia de un filtro (único) que separa en componentes independientes sin necesidad de imponer más que una serie de propiedades encontradas en los estímulos visuales -las simetrías estadísticas de las imágenes naturales-. Este filtro va a tener el soporte matemático de las funciones de wavelet por compartir éstas propiedades con los estímulos visuales. Por ello, haremos un breve repaso al análisis multiresolución de señales. Acabaremos mostrando cómo con un único filtro (aun siendo óptimo) no somos capaces de lograr una buena representación.

### 2.1. Imágenes naturales

Con imágenes naturales entendemos a todas aquellas imágenes que proceden de escenas del mundo natural. Éstas tienen soporte en 2D pero, como sabemos, están formadas por las proyecciones polares del escenario natural tridimensional. Desde un

punto de vista del procesamiento visual estas imágenes son los estímulos sensoriales y si uno conoce las propiedades de éstas podrá comprender el funcionamiento de su procesamiento.

Vamos a necesitar definir las señales de la siguiente forma: sea  $I(\vec{x})$  la intensidad de la imagen, medida en niveles de gris, en un punto (o píxel)  $\vec{x}$ . Por razones que expondremos a lo largo del desarrollo conviene trabajar con el contraste, i.e.  $c(\vec{x}) \equiv I(\vec{x}) - I_0$ , siendo  $I_0$  una constante que hace que  $c(\vec{x})$  tenga media nula sobre toda la imagen,

$$\int d\vec{x} c(\vec{x}) = 0. \quad (2.1.1)$$

Aunque ahora estemos haciendo un desarrollo teórico, conviene notar que las magnitudes anteriormente definidas estarán discretizadas de tal forma que podemos interpretar el contraste como una matriz de puntos, donde la posición de los puntos de la matriz estará relacionada con la posición física del contraste y cada uno de los puntos tomará valores (también en este caso discretos) que estarán comprendidos dentro del rango dinámico. Generalmente, el rango dinámico se le conoce como profundidad de gris<sup>1</sup> y viene medido en bits.

El contraste de las imágenes  $c(\vec{x})$  sigue un proceso estocástico. Esto es: una imagen natural es la manifestación de un suceso aleatorio que tendrá asociado una probabilidad de que ese suceso ocurra. Para definir completamente al espacio que componen las posibles imágenes naturales necesitamos conocer la distribución de probabilidades de la señal. Esta distribución es una función que a cada contraste le asocia una probabilidad:  $P[\{c(\vec{x})\}_{\vec{x}}]$ . Esta distribución de probabilidades es un concepto puramente teórico, de la que en modo alguno somos capaces de dar siquiera una buena estimación experimental. Para poder estudiar las propiedades de las imágenes naturales sin conocer su distribución deberemos tener un colectivo lo suficientemente representativo para que el uso de los estimadores estadísticos nos den una visión fiable de sus momentos asociados.

---

<sup>1</sup>Cuando se trabaja con imágenes en color, además de la profundidad de gris, para definir las propiedades de un píxel se necesita la profundidad de color

Es importante resaltar por último que cada ser vivo se nutrirá de diferentes estímulos visuales porque su realidad visual es diferente, con lo que a la hora de hablar de colectivos de imágenes naturales tendremos que tener en cuenta las diferentes dietas visuales de los distintos seres vivos .

## 2.2. Estadística de las imágenes naturales

Cuando uno mira al mundo que le rodea, podría pensar que las imágenes naturales que observa son objetos complicados, sin ninguna estructura subyacente. Podría también pensar que en una misma imagen, o entre imágenes diferentes, no existe alguna correlación. Si esto fuese así, el sistema visual no tendría más remedio que considerar que cada píxel se comporta de manera independiente.

Hagamos un sencillo cálculo para ilustrar lo que supone trabajar con imágenes que no poseen estructura. Tomemos imágenes de 16 por 16 píxeles con una profundidad de gris por píxel de 8 bits (256 niveles de gris). Si suponemos que no existe ningún tipo de correlación, es decir que todos los píxeles se comportan de manera independiente, entonces estaríamos ante un espacio probabilístico de  $2^{768}$  eventos, ¡habría más de  $1.5 \cdot 10^{231}$  posibilidades!.

No es cierto que los píxeles en las imágenes naturales se comporten de manera independiente por dos razones. Primero, porque el número de imágenes que podría almacenar el sistema visual sería mucho menor que el que estudios biológicos afirman que almacena. Segundo, porque existen ciertas correlaciones estadísticas -como vamos a ver en este capítulo- que permiten reducir a una fracción pequeña el espacio de las imágenes posibles.

Hay que insistir en que las simetrías estadísticas que vamos a describir a continuación son eso, estadísticas, por ello no conviene intentar extrapolar el análisis del conjunto estadístico a los individuos. Tampoco conviene olvidar las propias limitaciones que nos imponen nuestros aparatos de medida, como por ejemplo el tamaño o la resolución finita. Esto va a provocar que cuando queramos explorar nuestras imágenes a tamaños o escalas que estén cercanos a los límites impuestos, las simetrías no se

manifiesten. Ahora bien, lo que sí que podremos hacer aquí es extrapolar, para poder eliminar los errores espurios que introduce nuestro aparato de medida.

### 2.2.1. Invariancia traslacional

Esta simetría es la más simple de entender de todas y nos dice que no existe un centro del universo, no hay puntos privilegiados, y por tanto no se puede distinguir un píxel de otro. Dada una imagen, un objeto que aparece en un lugar determinado tendrá la misma probabilidad de aparecer en un lugar diferente en otra imagen. Por lo tanto, en las imágenes naturales nos encontraremos que, por ejemplo un árbol aparece a la izquierda de la escena con la misma frecuencia que aparece en cualquier otro lugar.

Expresando esta idea de forma matemática, tenemos que si  $P[\{c(\vec{x})\}_{\vec{x}}]$  es la distribución de probabilidades de la variable aleatoria  $c(\vec{x})$  entonces invariancia traslacional querrá decir,

$$P[\{c(\vec{x})\}_{\vec{x}}] = P[\{c(\vec{x} + \vec{x}_0)\}_{\vec{x}}] , \quad (2.2.1)$$

donde  $\vec{x}_0$  es cualquier vector.

### 2.2.2. Ergodicidad

Esta propiedad tiene alguna relación con el carácter fractal, que en la próxima sección discutiremos, de las imágenes. Si tomamos una imagen cualquiera con nuestro aparato de medida, estaremos prefijando el tamaño de ésta, y uno podría llegar a pensar que ese es el tamaño válido y único. Pero una imagen, a su vez, está compuesta por diferentes subimágenes (dependiendo de lo que uno haya tomado podrá distinguir: árboles, nubes, edificios ...) que tendrían la categoría de *imagen* si hubiésemos tomado como tamaño de partida el que abarca a esos objetos. Podríamos decir ahora que esas subimágenes están formadas a su vez por partes más pequeñas, y otorgar a estas últimas la misma categoría que se ganaron en la frase anterior las *subimágenes*. De la misma forma, podríamos decir que la primera imagen que tomamos con nuestro aparato de medida, no es más que una parte de una imagen más grande. Es por ello que concluimos

que no hay una escala fija. Por lo tanto, la ergodicidad en las imágenes naturales nos va a permitir que, si uno quiere tomar promedios de cantidades intensivas definidas en un colectivo de imágenes, podrá hacerlo o bien sobre el colectivo, o bien a lo largo de una sola imagen, puesto que el propio colectivo podría provenir de una sola imagen, así como la imagen podría dar lugar a un colectivo.

Expuesto con el rigor matemático de una ecuación,

$$\langle s_i(\vec{x}) \rangle_{\vec{x}, i=i_0} = \langle s_i(\vec{x}) \rangle_{\vec{x}=\vec{x}_0, i}, \quad (2.2.2)$$

siendo  $s$  cualquier variable intensiva definida en nuestro colectivo de imágenes,  $i$  es un índice que representa una imagen en el colectivo. y el subíndice 0 denota que se ha fijado o bien la imagen, o bien el píxel.

### 2.2.3. Invariancia de escala

La invariancia de escala es negar la existencia de una escala universal en las imágenes naturales. Esto implica que el tamaño de los objetos que aparecen en las escenas (que depende del tamaño natural del objeto y de la distancia a la que se encuentre el observador) no será fijo y tendrá una distribución de probabilidades independiente de escala.

Veámoslo expresado matemáticamente,

$$P[\{c(\vec{x})\}_{\vec{x}}] = P[\{r^\mu c(r\vec{x})\}_{\vec{x}}], \quad (2.2.3)$$

siendo  $r$  el factor con el que dilatamos (o contraemos) la imagen y  $\mu$  es la menos codimensión del soporte de  $c(\vec{x})$ . Ésta viene definida como la diferencia entre la dimensión del soporte de  $c(\vec{x})$  y la dimensión del espacio, por tanto siempre toma valores negativos (ver [17] para más información).

### Autosimilaridad y fractales

La invariancia de escala implica que las imágenes son autosimilares en sentido estadístico. Un objeto es autosimilar estadístico cuando se puede definir alguna variable



Figura 2.1: Curva de Koch. Ejemplo de un objeto autosimilar afín.

aleatoria en dicho objeto, y esta variable manifiesta alguna propiedad que de manera estadística se mantiene bajo cambios de escala. En el caso de las imágenes naturales, el objeto que es autosimilar estadístico es la imagen, el contraste es una variable aleatoria definida en las imágenes cuya distribución de probabilidades mantiene la misma estructura bajo cambios de escala.

Existen otros objetos que bajo un cambio de escala mantiene la misma forma o dicho de otra manera, tales que algunas de sus partes son versiones re-escaladas de ellos mismos. Estos objetos se les conoce como autosimilares afines. Esta propiedad es más estricta que la autosimilaridad en sentido estadístico, y un objeto que sea autosimilar afín lo será también en sentido estadístico (la implicación inversa, evidentemente, no es cierta). Ejemplos de estos objetos, los podemos ver en las figuras 2.1 y 2.2.

Los objetos autosimilares son fractales puesto que tienen *dimensión de Hausdorff*<sup>2</sup> no triviales.

#### 2.2.4. Espectro de potencias

Una propiedad muy interesante que es consecuencia de la autosimilaridad de las imágenes es el comportamiento del espectro de potencias  $S(\vec{f})$ . Veamos como podemos relacionarlo con la correlación a dos puntos del contraste  $\mathcal{C}(\vec{x}, \vec{y})$ . Definimos esta última

---

<sup>2</sup>Un fractal es un conjunto matemático que viene caracterizado por un número positivo que llamaremos dimensión, y que en los casos triviales coincide con la definición usual de dimensión euclídea. La *dimensión de Hausdorff* caracteriza las propiedades topológicas y funcionales del conjunto.



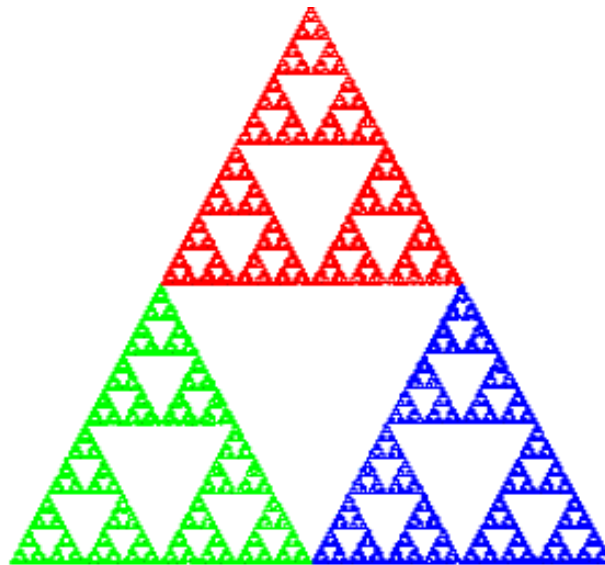


Figura 2.2: Triángulo de Sierpinski, objeto autosimilar afín. Cada uno de los tres triángulos con distintos colores que forman el triángulo de Sierpinski son iguales entre sí e iguales a este último.

como

$$\mathcal{C}(\vec{x}, \vec{y}) \equiv \langle c(\vec{x})c(\vec{y}) \rangle . \quad (2.2.4)$$

Debido a la invariancia traslacional la correlación no puede depender de los puntos  $\vec{x}$  e  $\vec{y}$  de forma independiente, si no que sólo puede depender de su diferencia  $\vec{x} - \vec{y} \equiv \vec{r}$ ,

$$\mathcal{C}(\vec{x}, \vec{y}) = \mathcal{C}(\vec{r}) . \quad (2.2.5)$$

Esto es muy interesante: al depender la correlación sólo de una variable, ésta, en el espacio de frecuencias -haciendo la transformada de Fourier en  $\vec{x}$  e  $\vec{y}$ -, contendrá la delta en una variable: la transformada de Fourier ha diagonalizado la correlación, es decir, en este espacio las imágenes están decorreladas. Por tanto, si las imágenes siguieran una estadística gaussiana -veremos luego que no la siguen- la transformada de Fourier conseguiría la independencia de estas imágenes en el espacio de frecuencias. Para que en el espacio directo las diferentes imágenes del colectivo fueran también independientes -siempre que siguieran esa estadística- habríamos de eliminar la dependencia frecuencial en la transformada de Fourier de la correlación.

Para evaluar la correlación hemos de hacer uso de un buen estimador al no conocer la distribución de probabilidades. El estimador natural del valor medio es el promedio o media aritmética, que habrá que hacer sobre todo el colectivo de imágenes. Si hacemos uso de la ergodicidad, promediaremos dentro de la imagen todos los puntos de ésta. Esto es muy significativo, nos permite identificar  $\mathcal{C}(\vec{r})$  como la autocorrelación del contraste. Debido a que la correlación no depende del punto hemos podido hacer uso de la ergodicidad.

El espectro de potencias  $S(\vec{f})$  se define como la transformada de Fourier de  $\mathcal{C}(\vec{r})$  respecto de  $\vec{r}$ . Entonces, la transformada de la autocorrelación será

$$S(\vec{f}) = |\hat{c}(\vec{f})|^2 . \quad (2.2.6)$$

Indicios experimentales muestran [8] que esta función presenta una ley de potencias para imágenes naturales en  $f \equiv |\vec{f}|$ , es decir que verifica



Figura 2.3: Imagen de Lena

$$S(\vec{f}) \sim \frac{1}{f^{2-\epsilon}}, \quad (2.2.7)$$

siendo  $\epsilon$  una cantidad pequeña y que dependerá del colectivo de imágenes que estemos estudiando (para más detalles ver [18], [19] y [8]).

Concluimos verificando lo que anunciábamos al comienzo: el comportamiento del espectro de potencias es consecuencia de la invariancia de escala, puesto que una ley de potencias indica, al menos en cuanto a correlación, que no existe escala fija. Hay que señalar que hay autores que arguyen que el comportamiento del espectro de potencias se debe a la presencia de bordes en las imágenes, con los bordes distribuidos de una forma invariante de escala.

### Decorrelación de una imagen

La transformada de Fourier del contraste es una cantidad compleja, por tanto



Figura 2.4: Imagen de Lena tras multiplicar por  $|\vec{f}|$  en el espacio de Fourier (aproximador de decorrelación).

$$\hat{c}(\vec{f}) = |\hat{c}(\vec{f})|e^{i\phi(\vec{f})}, \quad (2.2.8)$$

siendo  $\phi(\vec{f})$  la fase.

Hemos visto que  $|\hat{c}(\vec{f})|$  sigue una ley de potencias. Por tanto, si queremos decorrelar una imagen necesitamos eliminar la dependencia frecuencial en la correlación (en su transformada), una buena aproximación consistirá en multiplicar el módulo de la transformada de Fourier por el módulo de  $\vec{f}$ . La decorrelación es una estrategia usada por las células ganglionares del nervio óptico para eliminar redundancia, en [6] se argumenta que estas células decorrelan la señal para hacer más compacto el código.

En las figuras 2.3 y 2.4 se puede apreciar la imagen de Lena <sup>3</sup> y el resultado de decorrelar ésta.

### 2.2.5. Estadística no gaussiana

Las imágenes naturales no pueden ser descritas sólo con el espectro de potencias. A pesar de que antiguos artículos (por ejemplo [8], [6] o [7]) defendían que las imágenes estaban completamente definidas mediante una estadística de segundo orden, es decir se consideraba a las imágenes como simples fractales.

Pero veamos en una sencilla experiencia como eso no es así. En la figura 2.5 hemos eliminado de la imagen de Lena su fase. Es decir, hemos calculado la transformada de Fourier del contraste y para todas las frecuencias nos hemos quedado sólo su módulo, despreciando la fase. Dado que en el módulo está relacionado con la estadística de orden dos, al haber eliminado la fase nos hemos quedado sólo con la estadística gaussiana. Es notorio que en la figura 2.5 lo único que se observa es la verificación de la ley de potencia que rige el comportamiento del espectro de potencias.

Sin embargo, si lo que hacemos es eliminar la estadística de orden dos, dividiendo la transformada de Fourier del contraste por su módulo -por definición estamos deco-

---

<sup>3</sup>Lena fue la playmate del mes de noviembre de 1972 de la revista PLAYBOY, y su póster central fue usado en las primeras pruebas para el procesamiento y transferencia de la imagen digital sobre la red ARPANET (la precursora de Internet) y desde entonces se ha tomado como estándar para testear los algoritmos de compresión.

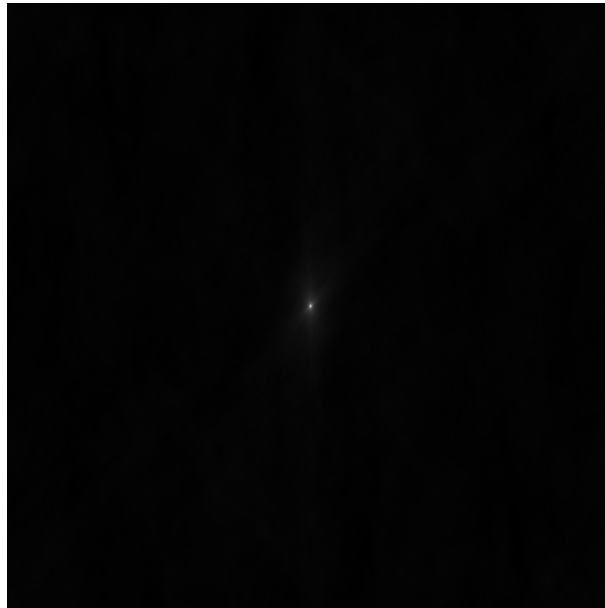


Figura 2.5: Imagen de Lena, habiendo eliminado la estadística de orden superior a dos.

rrelando, sin utilizar ninguna aproximación- la imagen- vemos en la figura 2.6 como todavía se observan mucha estructura de la imagen original. Comprobamos, una vez más, el comportamiento del espectro de potencias, al comparar esta última imagen decorrelada con la aproximación efectuada en la figura 2.4.

Por lo tanto concluimos que casi toda la información está contenida en la estadística de orden superior a dos.

### 2.2.6. Multiescala y multifractalidad

De los estudios de Ruderman [17] y Turiel et al. [20] y [21] se concluye que una completa descripción de la invariancia de escala en imágenes naturales requiere el análisis de la estadística de orden superior a dos.

Ruderman comprueba que las imágenes poseen propiedades de invariancia de escala, esto significa que en las imágenes naturales los objetos que aparecen están jerárquicamente estructurados en diferentes partes (bordes, barras, ...) siendo éstas invariantes

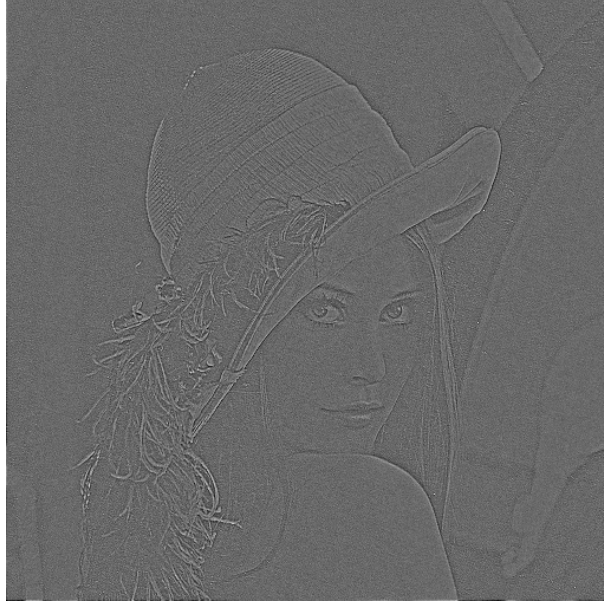


Figura 2.6: Imagen de Lena, eliminado la estadística de orden dos.

de escala, aunque con un comportamiento diferente bajo los cambios de escala. La transformación de cambio de escala no es la misma para cada una de las diferentes partes.

De manera matemática la multiescala se expresa a partir de los momentos de orden  $q$  de una variable aleatoria  $e_r(\vec{x})$  -que está relacionada con las diferentes partes que antes hablábamos- definida a una escala  $r$ . Debido a la invariancia de escala, estos momentos siguen una ley de potencias,

$$\langle e_r^q \rangle \sim r^{\gamma(q)}, \quad (2.2.9)$$

los exponentes (llamados exponentes autosimilares)  $\gamma(q)$  dependen de una manera no lineal con el orden  $q$  de los momentos. Si esta dependencia fuera lineal entonces estaríamos con objetos monofractales, y no necesitaríamos estadística de orden superior a dos.

Como resultado del estudio de la multiescala en las imágenes naturales ([20], [21], [22] y [23]) se llega al siguiente resultado crucial, que relaciona a dos escalas diferentes,

$r$  y  $L$  ( $L > r$ ), la variable estocástica  $e_r(\vec{x})$ ,

$$e_r \doteq \eta_{r,L} e_L , \quad (2.2.10)$$

por medio de una nueva variable aleatoria  $\eta_{r,L}$ , donde  $\doteq$  significa igualdad en términos de distribución, i.e. la igualdad no se cumple punto a punto, sino que los dos miembros de la ecuación tienen la misma distribución.

La nueva variable  $\eta_{r,L}$  sigue un proceso multiplicativo, lo que significa dos cosas,

1.  $\eta_{r,L}$  es independiente de  $e_L$
2. Es un proceso infinitamente divisible: dadas tres escalas  $L > r > s$  se satisface  $\eta_{s,L} \doteq \eta_{s,r} \eta_{r,L}$  ([24], [25]). Además, debido a la invariancia de escala, estas distribuciones sólo dependen del cociente entre escalas.

Esta propiedad estadística va a ser fundamental en todo el desarrollo de esta tesis. No nos equivocamos al decir que si las imágenes naturales no presentasen una estadística multiescala el presente estudio estaría vacío de contenido. Es nuestra piedra angular, como iremos viendo a lo largo de los capítulos. Conviene insistir, de igual manera, que el proceso multiplicativo que observamos en las imágenes naturales es, hasta los que hoy sabemos, una propiedad experimental u observacional, no teórica: no podemos deducir tan importante simetría a partir de primeros principios.

## 2.3. Análisis en wavelets

Toda función que se comporte de manera suficientemente suave la podemos descomponer en serie de Fourier, Laplace, etc. Esto es, podremos expandir la función como suma de unas ciertas funciones-base que irán multiplicando a unos coeficientes. Estos coeficientes se obtendrán por proyección de la función con cada una de las funciones base. La elección de la base dependerá de las características que queramos analizar pues los coeficientes de la expansión contendrán información que queramos extraer de nuestra función. Así, cuando hacemos una expansión en serie de Fourier de una función,



desarrollamos ésta en frecuencias por lo que cada coeficiente nos da idea del peso que tiene cada frecuencia en la función a estudiar.

En nuestro caso vamos a querer implementar las propiedades de invariancia traslacional y de escala en la base elegida. Esto nos permitirá hacer una representación donde esas propiedades estarán contenidas explícitamente, de tal forma que podamos extraer más fácilmente las componentes independientes. Hay que señalar que no existe una representación lineal que aglutine las dos invariancias anteriormente habladas; aún así es posible construir un código que sea invariante traslacional para cambios discretos de escala e invariante de escala para desplazamientos discretos [26].

Hay que dejar claro en este punto que nuestra función (en nuestro caso será la señal del contraste) posee una serie de límites, tiene tamaño y resolución finita. Esto nos va a permitir por un lado truncar la expansión, olvidándonos de los siempre incómodos restos teóricos. Pero, por otro lado, va a provocar la ruptura de las invariancias en los límites experimentales de la señal.

La expansión en wavelets discretas va a ser la candidata que mejor se adapte a nuestros propósitos. En este tipo de expansión la señal es descompuesta en diferentes niveles de detalles, del más grosero (escala integral) al más fino (píxel) a través de la base de wavelets. La base de wavelets se obtiene a partir de dilataciones y traslaciones -de esta forma se garantiza la invariancia de traslación y escala- de una *wavelet madre*, que es una función  $\Psi \in (L^1 \cap L^2)$  con media nula. Es por esto que necesitamos trabajar con el contraste en vez de la intensidad. Estas funciones típicamente no suelen tener muchos pasos por cero, es decir, que tienen pocas oscilaciones.

Antes de terminar esta sección veamos cómo se representa la proyección del contraste en una wavelet continua de escala  $l$  y posición  $\vec{x}$ ,

$$\alpha_{\Psi}(l, \vec{x}) \equiv \int d\vec{y} \, c(\vec{y}) \frac{1}{l^2} \Psi\left(\frac{\vec{x} - \vec{y}}{l}\right), \quad (2.3.1)$$

Esta proyección, la transformada de wavelet continua del contraste, representa un *zoom* de los detalles de la señal alrededor del punto  $\vec{x}$  a la escala  $l$ . Estos parámetros de escala y traslación varían de forma continua sobre  $\mathbb{R}$  y  $\mathbb{R}^2$ , respectivamente (con la limitación  $l > 0$ ).

### 2.3.1. Multiresolución

La transformada de wavelet continua de la señal es una función de 3 variables  $(l, \vec{x})$ , y a partir de ésta es posible reconstruir de manera exacta la señal original integrando en esas 3 variables a partir de la fórmula de *resolución de la identidad* [27]. En principio, cualquier función de  $L^2(\mathbb{R}^2)$  puede ser candidata a wavelet con tal de que cumpla la condición de admisibilidad (media nula). Sin embargo, qué sentido tiene hacer una transformación en la que pasamos de una señal (continua, en teoría) con dos variables a una con tres (también continua). No hemos ganado nada. Es por ello por lo que debemos utilizar la transformada de wavelet discreta: la información de la señal original podrá ser reducida a una representación mediante un número numerable de coeficientes.

Sabemos que las imágenes naturales tienen cierta redundancia -sobre todo entre escalas-, por ello, estamos interesados en una representación que haga uso de las propiedades estadísticas antes revisadas. Queremos de alguna forma *ordenar* de manera astuta nuestra información para luego poder deducir las componentes independientes que permitan codificar la señal de forma más compacta. Vamos a utilizar un tipo de representación en la que el subespacio generado por un nivel de resolución dado esté contenido en el subespacio generado por el siguiente nivel más grosero. En este esquema las wavelets,  $\Psi_{j\vec{k}}$ , constituyen un sistema de bases ortonormal en  $L^2(\mathbb{R}^2)$ . Este es el escenario elegido en el contexto de procesamiento y comprensión de imágenes naturales [28].

## 2.4. Expansión en una base ortonormal

Consideremos la proyección del contraste  $c(\vec{x})$  en una base diádica de wavelets  $\tilde{\Psi}_{j\vec{k}}(\vec{x}) = \tilde{\Psi}(2^j\vec{x} - \vec{k})$ . Dado que tenemos que usar wavelets discretas usamos la más simple de ellas, la diádica. Se llama así porque entre dos niveles de resolución hay un factor dos i.e., para pasar de una escala a otra más fina basta con dividir por dos. Cada wavelet  $\tilde{\Psi}_{j\vec{k}}$  es un proyector a una escala cuyo tamaño viene dado por  $2^{-j}$ , siendo  $j$  un número entero ( $j \in \mathbb{Z}$ ); y centrada en un punto  $\vec{x}_0 = 2^{-j}\vec{k}$ , donde el vector

$\vec{k} \equiv (k_1, k_2)$  con  $k_1, k_2 \in \mathcal{Z}$ . Si la base discreta es completa, entonces el contraste puede ser expandido en una base de wavelets ortogonal a  $\tilde{\Psi}$  usando sus proyecciones como coeficientes. La base dual  $\Psi_{j\vec{k}}$  debe satisfacer:

$$\langle \tilde{\Psi}_{j'\vec{k}'} | \Psi_{j\vec{k}} \rangle = 2^{-2j} \delta_{jj'} \delta_{\vec{k}\vec{k}'} , \quad (2.4.1)$$

donde el  $\langle \tilde{\Psi}_{j'\vec{k}'} | \Psi_{j\vec{k}} \rangle$  significa el producto escalar entre las funciones  $\tilde{\Psi}_{j'\vec{k}'}$  y  $\Psi_{j\vec{k}}$ ,

$$\langle \tilde{\Psi}_{j'\vec{k}'} | \Psi_{j\vec{k}} \rangle \equiv \int d\vec{x} \tilde{\Psi}_{j'\vec{k}'}(\vec{x}) \Psi_{j\vec{k}}(\vec{x}) . \quad (2.4.2)$$

Salvo que se diga explícitamente lo contrario a la hora de la reconstrucción de la señal asumiremos por simplicidad que estamos trabajando con una base de wavelets autodual ( $\tilde{\Psi}_{j'\vec{k}'} = \Psi_{j\vec{k}}$ ). Dado que las imágenes tienen un tamaño finito podemos tomar  $j \geq 0$  (olvidándonos de los ordenes más groseros, ver sección del apéndice A.2) donde  $\Psi_{0,\vec{0}}$  cubrirá nuestra imagen y será la wavelet madre,  $\Psi_{0,\vec{0}} \equiv \Psi$ . Por la misma razón el rango de  $\vec{k}$  vendrá acotado de la siguiente forma: a la escala  $j$  sólo puede tomar los valores  $k_1, k_2 = 0, 1, \dots, 2^j - 1$ . Hechas estas aclaraciones el contraste se podrá expandir de la siguiente forma,

$$c(\vec{x}) = \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathcal{Z}_{2^j})^2} \alpha_{j\vec{k}} \Psi_{j\vec{k}}(\vec{x}) , \quad (2.4.3)$$

donde los coeficientes  $\alpha_{j\vec{k}}$  vienen dados de particularizar la ecuación 2.3.1 a una base diádica,

$$\alpha_{j\vec{k}} \equiv 2^{2j} \langle \tilde{\Psi}_{j\vec{k}} | c \rangle = 2^{2j} \int d\vec{x} \tilde{\Psi}(2^j \vec{x} - \vec{k}) c(\vec{x}) . \quad (2.4.4)$$

La ecuación 2.4.3 puede ser interpretada de dos maneras: como una fórmula para reconstruir  $c(\vec{x})$  una vez que sabemos los coeficientes  $\alpha_{j\vec{k}}$ , o bien, otra manera de escribir  $c(\vec{x})$  como superposición de las distintas wavelets,  $\Psi_{j\vec{k}}(\vec{x})$ .

Para cualquier wavelet que usemos, los coeficientes  $\alpha$ 's deben seguir el proceso multiplicativo, consecuencia del carácter multiescala de las imágenes. Con lo que la ecuación 2.2.10 para nuestros coeficientes toma la forma

$$\alpha_{j\vec{k}} \doteq \eta_{j\vec{k}} \alpha_{j-1[\frac{\vec{k}}{2}]} . \quad (2.4.5)$$

Recordemos que las variables estocásticas  $\eta_{j\vec{k}}$  son independientes de  $\alpha_{j-1[\frac{\vec{k}}{2}]}$  y debido a que el cociente entre escalas está fijado a dos, tienen la misma distribución para todas los niveles de resolución  $j$  y todas la posiciones espaciales  $\vec{k}$ .

## 2.5. Fórmula de la wavelet óptima

Si en la ecuación con la que terminábamos la sección anterior, imponemos que se cumpla punto a punto, es decir que sea una igualdad estricta para toda imagen, resolución y posición, esto implicará unicidad, i.e. si existe una wavelet que cumple tal relación será única. A partir de esta nueva relación (no es más que la ecuación 2.4.5 sin punto) podemos despejar los  $\eta_{j\vec{k}}$ ,

$$\eta_{j\vec{k}} = \frac{\alpha_{j\vec{k}}}{\alpha_{j-1[\frac{\vec{k}}{2}]}}. \quad (2.5.1)$$

Estas variables  $\eta_{j\vec{k}}$  van a definir una representación en la que cada nivel de resolución va a ser independiente respecto de otro. Los  $\eta_{j\vec{k}}$  y  $\eta_{j'\vec{k}'}$  serán estadísticamente independientes para  $j \neq j'$  [29]. Es por ello que llamamos a esta representación óptima.

Nuestro propósito ahora va a ser intentar extraer a partir de la ecuación 2.4.3 la wavelet óptima. Partiendo de la relación multiplicativa que relaciona los  $\alpha$ 's y los  $\eta$ 's es fácil llegar a la relación

$$\alpha_{j\vec{k}} = \prod_{i=0}^{j-1} \eta_{j-i[\frac{\vec{k}}{2^i}]} \alpha_{0\vec{0}}. \quad (2.5.2)$$

Esta relación es bastante significativa, pues habiendo hecho uso de la wavelet óptima tenemos que todos los coeficientes se pueden deducir a partir del coeficiente a la escala más grosera,  $\alpha_{0\vec{0}}$ . Es decir, todos los coeficientes tienen en común un *coeficiente padre* que no depende de los índices  $j$  y  $\vec{k}$  del *hijo*. Para otro tipo de wavelets, la relación anterior sí que dependería de la escala y posición que estuviésemos mirando, y por tanto no habría un *coeficiente padre* común a todos los coeficientes.

Si ahora tomamos el promedio en el colectivo de imágenes de la ecuación 2.4.3, hacemos uso del resultado anterior, y tenemos en cuenta que los  $\eta_{j\vec{k}}$  tienen la misma

distribución (en verdad, sólo necesitamos que tengan el mismo valor para el momento de orden 1) y son independientes entre escalas diferentes tendremos

$$\langle c(\vec{x}) \rangle = \langle \alpha_{0\vec{0}} \rangle \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathbb{Z}_{2^j})^2} \langle \eta \rangle^j \Psi_{j\vec{k}}(\vec{x}) . \quad (2.5.3)$$

A partir de esta última relación podemos despejar  $\Psi$ , siempre bajo las hipótesis comentadas en la sección del apéndice A.2,

$$\Psi(\vec{x}) = \frac{1}{\langle \alpha_{0\vec{0}} \rangle} \left( \langle c(\vec{x}) \rangle - \langle |\eta| \rangle \sum_{\vec{k} \in (\mathbb{Z}_2)^2} \langle c(2\vec{x} - \vec{k}) \rangle \right) . \quad (2.5.4)$$

En vez de  $\eta$  hemos puesto sólo su valor absoluto al hacer dos hipótesis. Primera consideramos la independencia estadística entre el valor absoluto de  $\eta$  y su signo. Segundo, el valor esperado de los signos por invariancia de traslación debería ser cero, por lo tanto el promedio de los signos vendrá gobernado por el segundo término, es decir las fluctuaciones. Considerando que estas fluctuaciones son pequeñas e independientes de la escala y la posición, podemos interpretar entonces a las fluctuaciones como un factor global. En nuestro contexto  $\langle |\eta| \rangle$  vale  $\frac{1}{2}$  debido a la invariancia traslacional y de escala [22]. Esta fórmula no suele venir expresada de esta forma tan simple en la literatura. Originalmente la fórmula de obtención de la wavelet fue deducida en el espacio de Fourier y ésta toma la forma

$$\hat{\Psi}(\vec{f}) = \frac{1}{\langle \alpha_{0\vec{0}} \rangle} \left( \langle \hat{c}(\vec{f}) \rangle - \frac{\langle |\eta| \rangle}{4} \frac{\Lambda(\vec{f})}{\Lambda(\frac{\vec{f}}{2})} \langle \hat{c}(\frac{\vec{f}}{2}) \rangle \right) , \quad (2.5.5)$$

donde  $\Lambda(\vec{f}) \equiv (1 - e^{-2\pi f_1})(1 - e^{-2\pi f_2})$ .

Dado un colectivo de imágenes, suficientemente representativo, la wavelet queda completamente definida salvo por un factor multiplicativo; éste no nos importa lo que valga puesto que si  $\Psi$  es una wavelet válida,  $a\Psi$  (para todo  $a$  real distinto de cero) también lo es. La indeterminación de la constante se resuelve requiriendo que  $\Psi$  esté normalizada en norma 2, es decir

$$\langle \Psi | \Psi \rangle \equiv \int d\vec{x} \Psi^2(\vec{x}) = 1 . \quad (2.5.6)$$

### 2.5.1. Colectivo de imágenes utilizado

El colectivo consistió en 4212 imágenes obtenidas con una cámara fotográfica digital de Kodak modelo DCS420 (con una lente de 28 mm) y que forma la base de datos de Hans van Hateren ([30]). Las imágenes tienen una resolución de  $1536 \times 1024$  y una profundidad de gris efectiva de 12 bits. La resolución angular de las imágenes es de aproximadamente 1 minuto de arco por píxel. Los parámetros tales como el número ISO, la apertura del diafragma o la velocidad de disparo que regulan la intensidad luminosa fueron ajustados por el procesador de la cámara para cada imagen y registrados para su posterior consulta en [http://hlab.phys.rug.nl/imlib/im\\_list.txt](http://hlab.phys.rug.nl/imlib/im_list.txt). El colectivo de imágenes fue tomado en diferentes ambientes (bosque, áreas urbanas, espacios abiertos) para conseguir una gran variabilidad en los datos, y en las escenas se ven edificios, árboles, troncos, hojas, flores, etc.

## 2.6. Obtención de la wavelet óptima

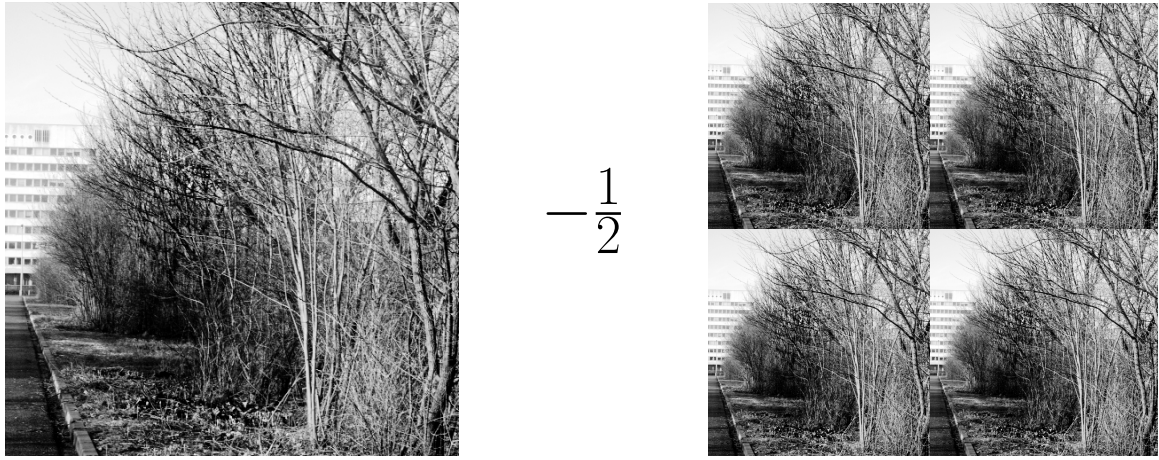
Dado el carácter lineal de la ecuación 2.5.4 para calcular la wavelet de un colectivo de imágenes, podemos o bien primero obtener el promedio del contraste sobre el colectivo de imágenes y aplicar la ecuación 2.5.4, o bien aplicar la fórmula para cada imagen, obteniendo una wavelet, no óptima, para cada imagen, y luego promediar las wavelets.

La gran ventaja de expresar la fórmula de la wavelet óptima en el espacio directo respecto al espacio de Fourier, es que nos permite entender mejor cómo se obtiene la wavelet. En la figura 2.7 vemos cómo se construye la wavelet para una imagen.

Si ahora cogemos un colectivo de 4167 imágenes (para evitar efectos de borde nos quedamos con la parte central:  $1024 \times 1024$ ), pertenecientes a la base de datos de Hans van Hateren [30] (fueron descartadas un total de 45 imágenes que estaban corrompidas), el resultado será el que se muestra en la figura 2.8

Si en vez de utilizar la fórmula 2.5.4 que opera en el espacio directo, usamos la ecuación 2.5, que trabaja en el espacio de Fourier, aplicada sobre el mismo colectivo de antes, obtenemos la wavelet representada en la figura 2.9.

Vemos como las dos wavelets obtenidas no se parecen todo lo que cabía esperar. Hay



$$c(\vec{x})$$

$$\sum_{\vec{k} \in (\mathbb{Z}_2)^2} c(2\vec{x} - \vec{k})$$

=



Wavelet individual de imk00040.imc

Figura 2.7: Construcción de la wavelet individual a partir de la imagen *imk00040.imc* de la base de datos de Hans van Hateren (ver [30] para detalles sobre la imagen)

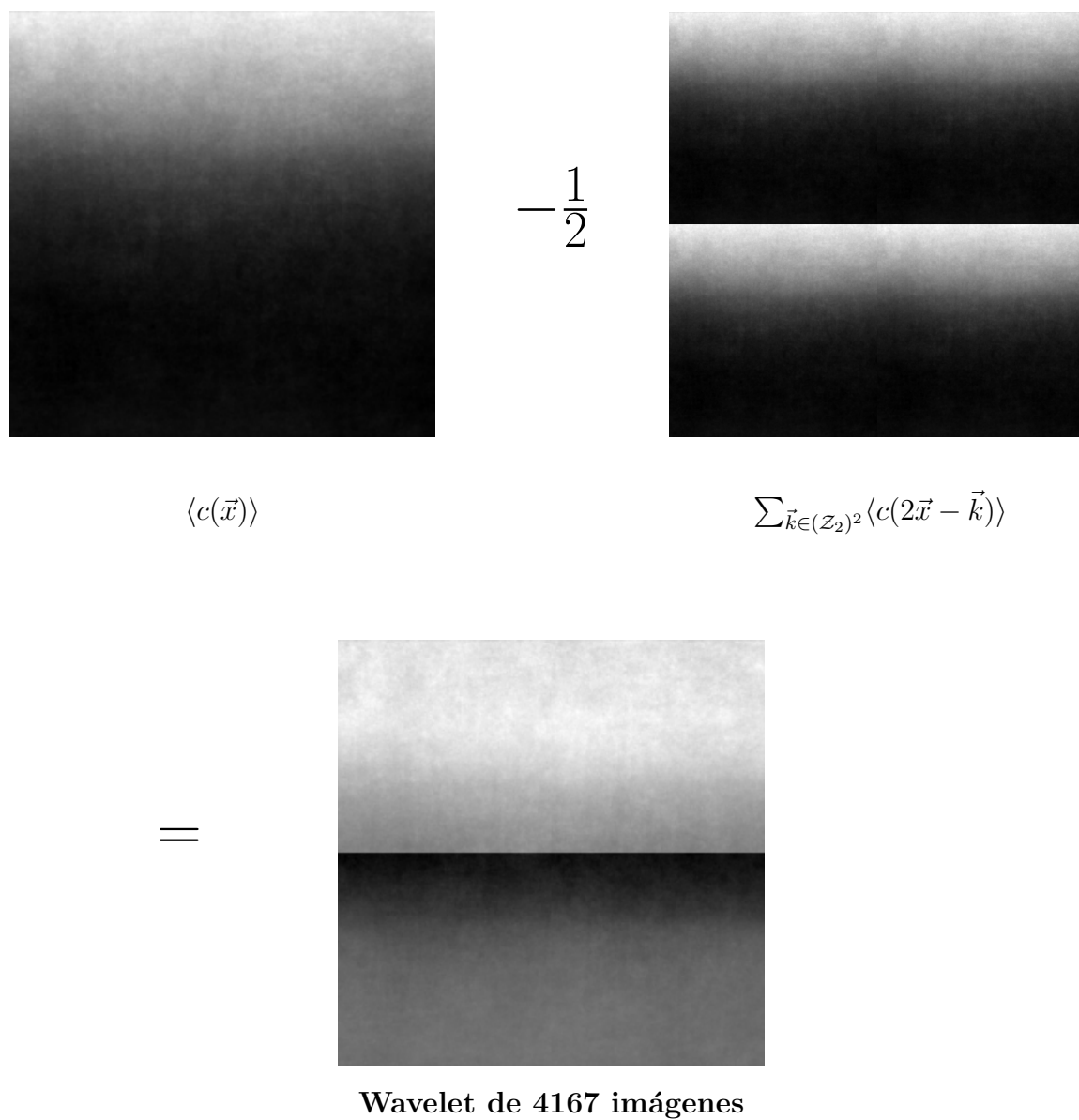


Figura 2.8: Construcción de la wavelet a partir de un colectivo de 4167 imágenes de la base de datos de Hans van Hateren [30]



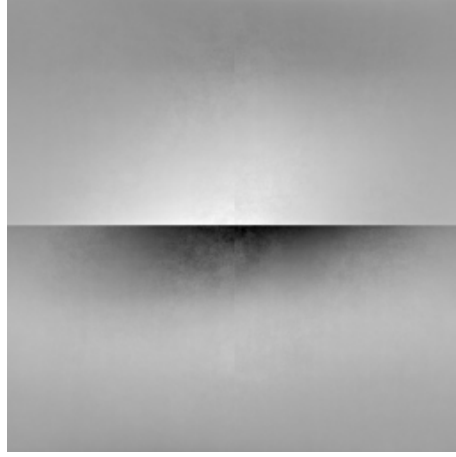


Figura 2.9: Wavelet obtenida a partir de la fórmula 2.5 sobre un colectivo de 4167 imágenes de la base de datos de Hans van Hateren [30].

que darse cuenta de que es la misma fórmula pero representada en dos espacios distintos. La respuesta reside en la interpretación que hemos hecho del segundo término en las dos fórmulas. Esta discusión se encuentra en la sección del apéndice A.3 (se recomienda leer antes los capítulos 3 y 4).

En ([13]) se observó que la wavelet, obtenida mediante el espacio de Fourier, presenta a grandes rasgos una simetría izquierda-derecha y una antisimetría arriba-abajo. Finalmente, implementando estas dos simetrías obtenemos la wavelet que representamos en la figura 2.10, y los cortes de esta wavelet en la figura 2.11.

### 2.6.1. Validación del proceso multiplicativo

En la figura 2.12 ([31]) se muestra la evidencia numérica del proceso multiplicativo, 2.4.5. La existencia del proceso multiplicativo es *extremadamente robusta*; en particular es válida para cualquier wavelet (cuando imponemos que se cumpla *punto por punto* tenemos la wavelet óptima), [22]. En el gráfico izquierdo de la figura 2.12 se muestra el histograma de los coeficientes de wavelet  $\alpha_{j,k}$  a dos escalas consecutivas utilizando la wavelet de Haar y la predicción utilizando el proceso multiplicativo,

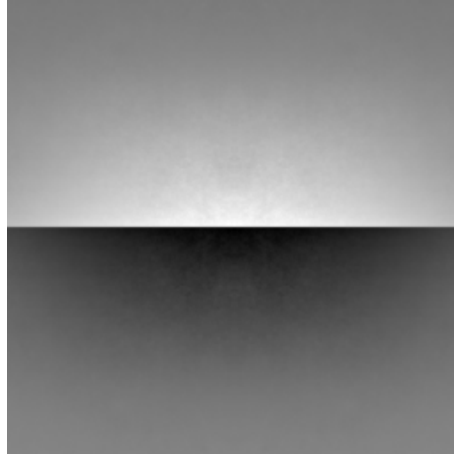


Figura 2.10: Wavelet obtenida a partir de la figura 2.9 implementando sus simetrías.

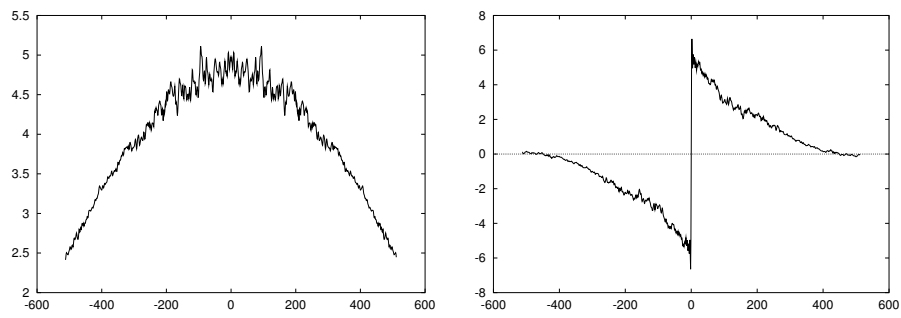


Figura 2.11: Cortes, horizontal y vertical, de la wavelet de la figura 2.10

mientras que en el gráfico derecho de la misma figura los cálculos anteriores han sido realizados con la wavelet óptima.

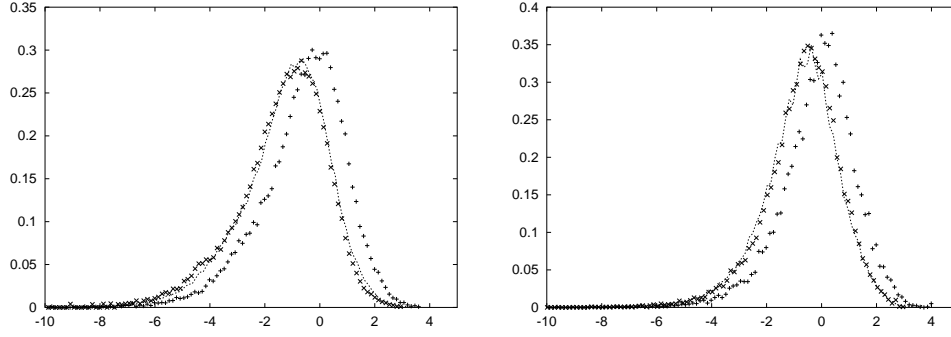


Figura 2.12: Verificación experimental del proceso multiplicativo, (2.4.5); comparación entre dos wavelets. **Izquierda:** Wavelet de Haar; **Derecha:** Wavelet óptima obtenida a partir de las 4167 imágenes, (2.10). Para el estudio se han utilizado un colectivo de 10 imágenes de la base de van Hateren. Empezando por el histograma (+) de los coeficientes de wavelet  $\alpha_{j,k}$  a la escala  $j = 5$ , asumiendo invariancia de traslación, y usando una distribución log-Poisson [21] de parametros:  $\Delta = 0.33$ ,  $\beta = 0.66$  y  $s = 1$  [22], obtenemos una predicción para la distribución de los  $\alpha_{j,k}$  a la escala  $j = 6$  (línea punteada). Debemos compararlo con la obtención directa de la distribución de los  $\alpha_{j,k}$  a la escala  $j = 6$  (x).

## 2.7. Persistencia: redundancia entre escalas y componentes independientes

Una vez obtenida la wavelet óptima vamos a ver las propiedades que presentan los coeficientes  $\alpha$  y  $\eta$  extraídos de diferentes imágenes utilizando esta wavelet.

En la figura 2.13 tenemos una imagen cogida del colectivo de van Hateren junto con tres representaciones de esta imagen. La wavelet utilizada ha sido la derivada de la gaussiana en la dirección horizontal. Estas representaciones sólo incluyen una escala

cada una, es decir en la ecuación 2.4.3 sólo sumamos a todos los  $\vec{k}$ , manteniendo fijo el índice  $j$ ,

$$\sum_{\vec{k} \in (\mathcal{Z}_{2^j})^2} \alpha_{j\vec{k}} \Psi_{j\vec{k}}(\vec{x}) . \quad (2.7.1)$$

Las escalas en este ejemplo no vienen identificadas por el índice  $j$  sino por el tamaño de la wavelet que se ha utilizado para construir las diferentes representaciones. De modo que, una wavelet del tamaño de un píxel corresponderá con la escala de más resolución de la imagen, si el tamaño de la wavelet es el doble que el de la anterior, estaremos ante la siguiente escala de más resolución, y así sucesivamente.

Un código compuesto por los coeficientes  $\alpha$  es lineal y redundante. Lineal por la propia definición de los coeficientes. El porqué es redundante requiere una mayor explicación. Parte del contenido de la información que tienen los coeficientes  $\alpha$  a una escala dada está también contenido en otras escalas. En la figura 2.13 vemos la persistencia de determinados objetos de una escala a otra. Con persistencia de objetos nos estamos refiriendo a lo siguiente: si a una determinada escala el código, formado por coeficientes  $\alpha$ , detecta (es decir, la amplitud de  $\alpha$  es grande) una determinada característica de la imagen (por ejemplo un borde) esta característica persistirá en las todas las escalas siguientes (más finas). Una vez detectado un detalle de una imagen, este persistirá en las sucesivas escalas. Esta persistencia, o redundancia, de los coeficientes  $\alpha$  hace que el código formado por estos coeficientes sea ineficiente para una codificación.

Analicemos el código formado por los coeficientes  $\eta$ . Como en el caso de los  $\alpha$  en la figura 2.14 representamos una imagen y la representación de los coeficientes  $\eta$  a tres escalas distintas, en cada de las tres representaciones está fijado el índice  $j$  y solo sumamos a todos los  $\vec{k}$

$$\sum_{\vec{k} \in (\mathcal{Z}_{2^j})^2} \eta_{j\vec{k}} \Psi_{j\vec{k}}(\vec{x}) . \quad (2.7.2)$$

El código  $\eta$  no es lineal, por ello en esta representación de la figura 2.14, que sí lo es, que hemos usado para analizar las propiedades de estos coeficientes no podemos visualizar representación alguna de la imagen. Aparte de la no linealidad del código  $\eta$  en esta figura apreciamos otras propiedades:

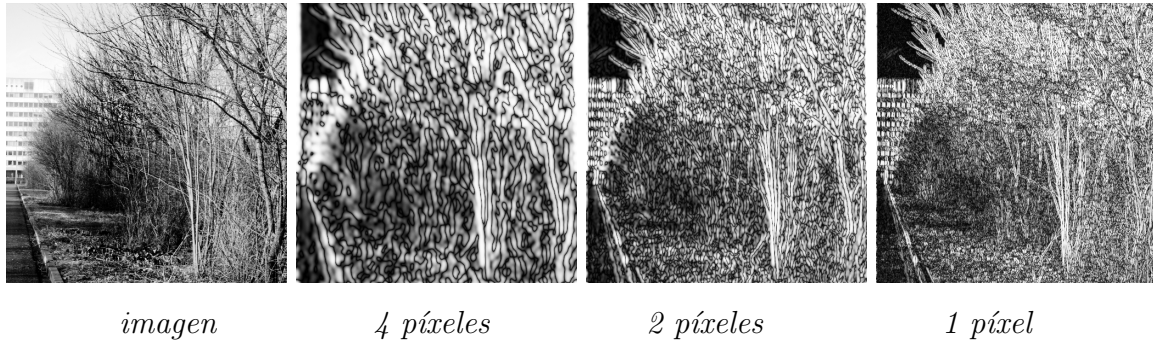


Figura 2.13: Imagen *imk00040.imc* y tres representaciones de ésta utilizando los coeficientes  $\alpha$  a tres escalas diferentes.

1. Independencia con el código construido a partir de los  $\alpha$ . Si comparamos las dos representaciones de los dos códigos podremos verificar que los coeficientes  $\eta$  son independientes de los  $\alpha$ .
2. La distribución es muy esparsa. Si nos fijamos en la figura 2.14 en una representación a una determinada escala, observamos que los valores que toma  $\eta$  son siempre cercanos a cero (valor representado por el color gris) excepto en unas pocas ocasiones que toma un valor distinto de cero, (estos valores se identifican con la wavelet a partir de la ecuación 2.7.2). Son esos valores los que van a almacenar la información relevante de las imágenes. Esta propiedad en la distribución de los  $\eta$  es consecuencia directa de tener un proceso multiplicativo (los procesos multiplicativos tienen siempre elevada curtosis [32]). Así como en las técnicas que aplican Código Esperso imponen que las distribuciones sean esparsas, en este modelo no se impone ad hoc, deduciéndose de las propiedades de las imágenes.
3. Independencia entre escalas. Si observamos dos escalas de la figura 2.14 notamos que no tienen nada en común. Con este código, pasar de una escala a otra implica un cierto grado de novedad. El código formado por los coeficientes  $\eta$  es independiente entre escalas, por tanto, es eficiente para una codificación.

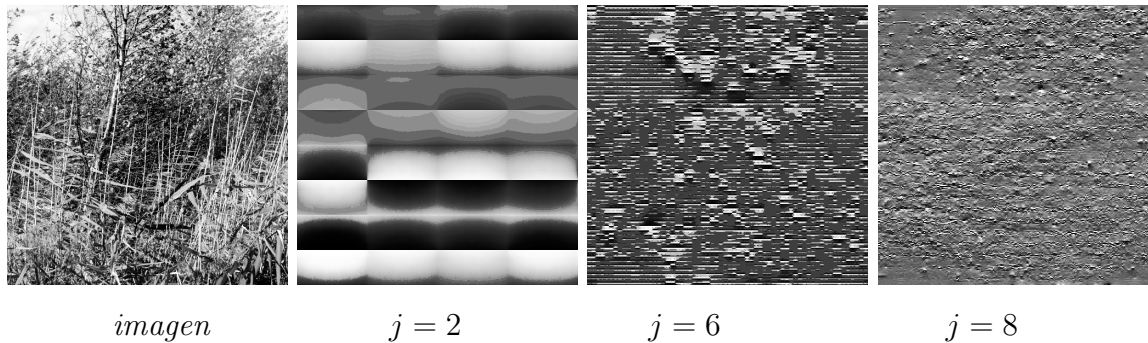


Figura 2.14: Imagen *imk03236.imc* y tres representaciones de los los coeficientes  $\eta$  a tres escalas diferentes.

## 2.8. Necesidad de más voces

La wavelet obtenida a partir del coléctivo de las 4167 imágenes va a contener la característica visual *media* que tienen en común estas imágenes. Sobre este asunto vamos a escribir largo y tendido a lo largo de los siguientes capítulos, ahora sólo es necesario interpretar esta característica visual *media* como el elemento común que tienen todas las imágenes del colectivo, de hecho, por la propia obtención de la wavelet, ésta -la característica visual- está asociada al promedio de todas las imágenes. Podemos considerar que la wavelet obtenida es un filtro detector de *ese* borde horizontal. Pero, ¿esto va ser suficiente para poder reconstruir un individuo del colectivo? ¿Nos servirá, esta wavelet así obtenida, para representar imágenes que no estén en dicho colectivo? Para responder estas dos interesantes preguntas qué mejor que verlo con dos ejemplos bastante ilustrativos. En la figura 2.15 vemos representaciones a distintas escalas de dos imágenes usando la wavelet óptima: una, *imk03236.imc* de la figura 2.14, perteneciente al catálogo; y otra, la ya conocida imagen de Lena, que no pertenece al colectivo de 4167 imágenes usadas para la obtención de la wavelet.

En las dos imágenes observamos como a medida que incluimos más escalas la representación es mejor, pero vemos que no es suficiente: la representación que incluye el mayor número de éscalas no es eficiente [13] para ninguna de las dos imágenes. Es más evidente para la imagen *imk03236.imc*, a pesar de pertenecer al colectivo de imágenes

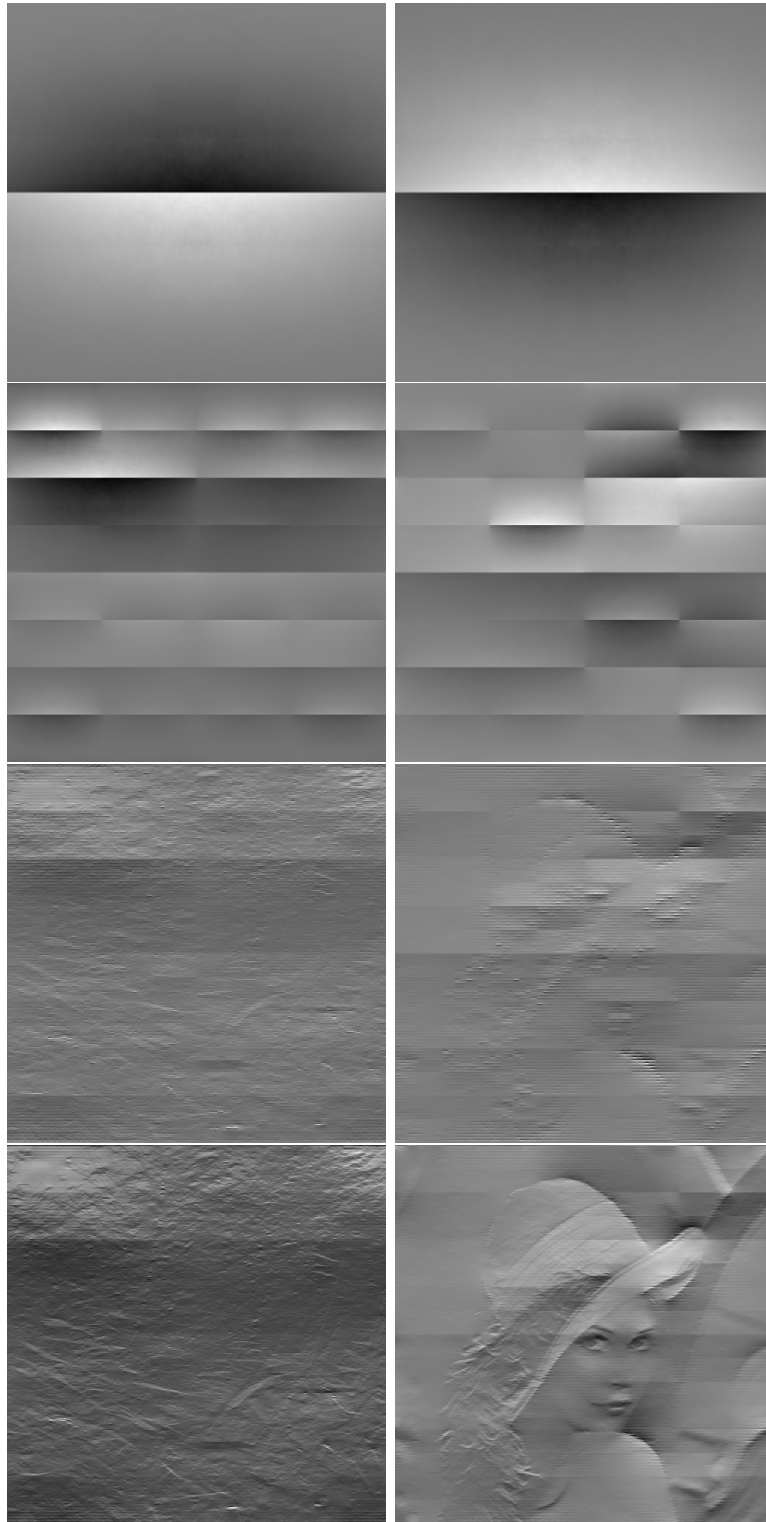


Figura 2.15:  $\sum_{j\vec{k}} \alpha_{j\vec{k}} \Psi_{j\vec{k}}(\vec{x})$  para  $j = 0, j \leq 2, j \leq 6$  y  $j \leq 8$  para las imágenes *imk03236.imc* y de Lena

usadas, por estar dominada por estructuras verticales. Concluimos, por tanto, que una única wavelet óptima no es suficiente para representar de forma eficiente las imágenes.

Deduzcamos cómo sería la fórmula de la wavelet óptima, entendida ahora ésta como el conjunto de todas sus diversas voces, incluyendo este nuevo grado de libertad. Necesitamos reescribir la expansión de una imagen en wavelets añadiendo un nuevo sumatorio: el de las voces. Entonces la ecuación 2.4.3 queda de la forma

$$c(\vec{x}) = \sum_{r=0}^{n-1} \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathbb{Z}_{2^j})^2} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x}) , \quad (2.8.1)$$

el índice  $r$  recorre las diferentes voces de wavelet. Un mero análisis algebraico muestra que, con bases separables, hacen falta  $l^d - 1$  elementos para formar una base completa, donde  $l$  es el paso de escala ( $l = 2$  para bases diádicas) y  $d$  es la dimensión del espacio ( $d = 2$  para imágenes). Por tanto, en nuestro modelo, harían falta 3 voces para formar una base completa. Para cada voz se cumple el proceso multiplicativo

$$\alpha_{rj\vec{k}} \doteq \eta_{rj\vec{k}} \alpha_{rj-1[\frac{\vec{k}}{2}]} , \quad (2.8.2)$$

si suponemos que existe la wavelet óptima (igualdad sin punto), está tendrá unos coeficientes  $\alpha_{rj\vec{k}}$  cuyo valor esperado para cada voz seguirá la siguiente relación (en verdad, para cualquier wavelet se cumplirá la igualdad, pero solo si es la óptima entonces también se cumplirá de forma exacta para los estimadores de estos valores esperados: las medias aritméticas)

$$\langle \alpha_{rj\vec{k}} \rangle = \langle \alpha_{r0\vec{0}} \rangle \langle \eta_r \rangle^j , \quad (2.8.3)$$

donde hemos tenido en cuenta las propiedades estadísticas de las variables  $\eta$  que necesitamos: independencia entre escalas e igualdad en la esperanza matemática (además de para todas las escalas habrá que asumir para todas las voces). A partir de esta última ecuación es fácil deducir que el valor esperado de  $c(\vec{x})$  toma la forma

$$\langle c(\vec{x}) \rangle = \langle \alpha_{00\vec{0}} \rangle \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathbb{Z}_{2^j})^2} \langle \eta \rangle^j \sum_{r=0}^{n-1} \frac{\langle \alpha_{r0\vec{0}} \rangle}{\langle \alpha_{00\vec{0}} \rangle} \Psi_{rj\vec{k}}(\vec{x}) , \quad (2.8.4)$$



siendo ahora

$$\langle \alpha_{0\vec{0}} \rangle \equiv \sum_{r=0}^{n-1} \langle \alpha_{r0\vec{0}} \rangle . \quad (2.8.5)$$

Definiendo la wavelet mezcla

$$\Phi_{j\vec{k}}(\vec{x}) \equiv \sum_{r=0}^{n-1} \frac{\langle \alpha_{r0\vec{0}} \rangle}{\langle \alpha_{0\vec{0}} \rangle} \Psi_{rj\vec{k}}(\vec{x}) , \quad (2.8.6)$$

llegamos al mismo resultado que obtuvimos para una única voz, ecuación 2.5.4, pero substituyendo  $\Psi$  por  $\Phi$ . Hay que interpretar  $\frac{\langle \alpha_{r0\vec{0}} \rangle}{\langle \alpha_{0\vec{0}} \rangle}$  como los pesos o contribuciones de cada voz. En principio son desconocidos y habrá que hacer algún tipo de hipótesis sobre el colectivo para poder deducirlos.

En nuestro contexto  $\langle |\eta| \rangle$  sigue valiendo  $\frac{1}{2}$  debido a la invariancia traslacional y de escala [22].

Hay que hacer notar que aunque una imagen individual no se pudo reconstruir con una sola wavelet, la imagen promedio sí se pudo, al ser ésta -la wavelet que utilizamos para reconstruir la imagen promedio- un promedio de las diferentes voces.

### 2.8.1. Bases orientadas

Motivados por la separación de las distintas voces de la wavelet óptima en [13] se propuso -evocados por la base de Haar en 2D- que las diferentes componentes de la base eran rotaciones de una de ellas.

El problema de separación de las distintas componentes está subdeterminado (más incognitas que ecuaciones), se ha de hacer alguna hipótesis sobre la forma de las voces. En la derivación del nuevo esquema se asumió que la wavelet obtenida anteriormente era una combinación de las distintas componentes a extraer, ecuación 2.8.6; que a su vez serán nuevos filtros detectores de características visuales; en verdad de la misma, pero versiones rotadas. Como ya hemos avanzado, se asumió que las distintas voces de la base se construían rotando una de ellas y que todas las voces eran ortogonales; si bien estas condiciones (versiones rotadas y ortogonalidad) hemos visto anteriormente que no son necesarias para su conexión con la wavelet *media*, son necesarias para la separación de las distintas componentes.

Para la separación de las distintas componentes de la base y uso del esquema rotacional se hizo, como primera aproximación, la hipótesis de que la wavelet mezcla obtenida sólo estaba formada por una componente de la nueva base,  $\Psi_0 = \Phi$ , y, por tanto, las otras voces se construirían simplemente rotando la wavelet conocida. Experimentalmente se observó que la aproximación realizada tenía una gran precisión.

En las figuras 2.16 y 2.17 se muestran los resultados tomando la base de wavelets con dos voces: la wavelet antigua y una versión rotada  $90^\circ$  de ésta, y con tres voces:  $\Phi$ , y dos versiones rotadas  $60^\circ$  y  $120^\circ$ . Las imágenes que se usaron fueron las mismas que en la figura 2.15. Podemos notar ya de forma cualitativa como la reconstrucción con dos voces es algo mejor que con tres, y, por supuesto, que con una única orientación. En [13] definieron el parámetro de calidad,

$$\epsilon_n[c] \equiv \sqrt{\frac{\sum_{r=0}^{n-1} \sum_{j\vec{k}} 2^{-2j} (\alpha_{rj\vec{k}})^2}{\langle c^2 \rangle}}, \quad (2.8.7)$$

que cuantificaba la eficiencia:  $\epsilon_n[c]$  próximo a 1 significa una buena representación de la expansión en wavelets. En la tabla 2.1 tenemos los resultados para las tres bases utilizadas: con 1, 2 ó 3 componentes/orientaciones.

La falta de ortogonalidad hace que  $\epsilon_n[c]$  pueda ser mayor que 1. Justamente, para poder cuantificar la ortogonalidad entre las distintas componentes con la wavelet  $\Phi$  a distintas escalas se definió la cantidad:

$$\epsilon_{rj} \equiv \sum_{\vec{k}} |\langle \Phi | \Psi_{rj\vec{k}} \rangle|. \quad (2.8.8)$$

En ésta se promedia en cada escala de las distintas posiciones  $\vec{k}$ . Los valores para las tres orientaciones están en la tabla 2.2. Estos deberían ser exactamente 0 para todo  $r$  y  $j$ , excepto para  $r = 1$  y  $j = 0$ :  $\epsilon_{10} = |\langle \Phi | \Phi \rangle| = 1$ . Se observó que la componente  $r = 2$  estaba muy cercana a la ortogonalidad con  $\Phi$ , sin embargo, para la componente  $r = 3$  se vió un pequeño acople a distintas escalas. Con todo, el solape más importante de, en torno al 10 %, se da entre las dos primeras escalas para la propia wavelet  $\Phi$ , atribuible a errores en la determinación de  $\Phi$ .

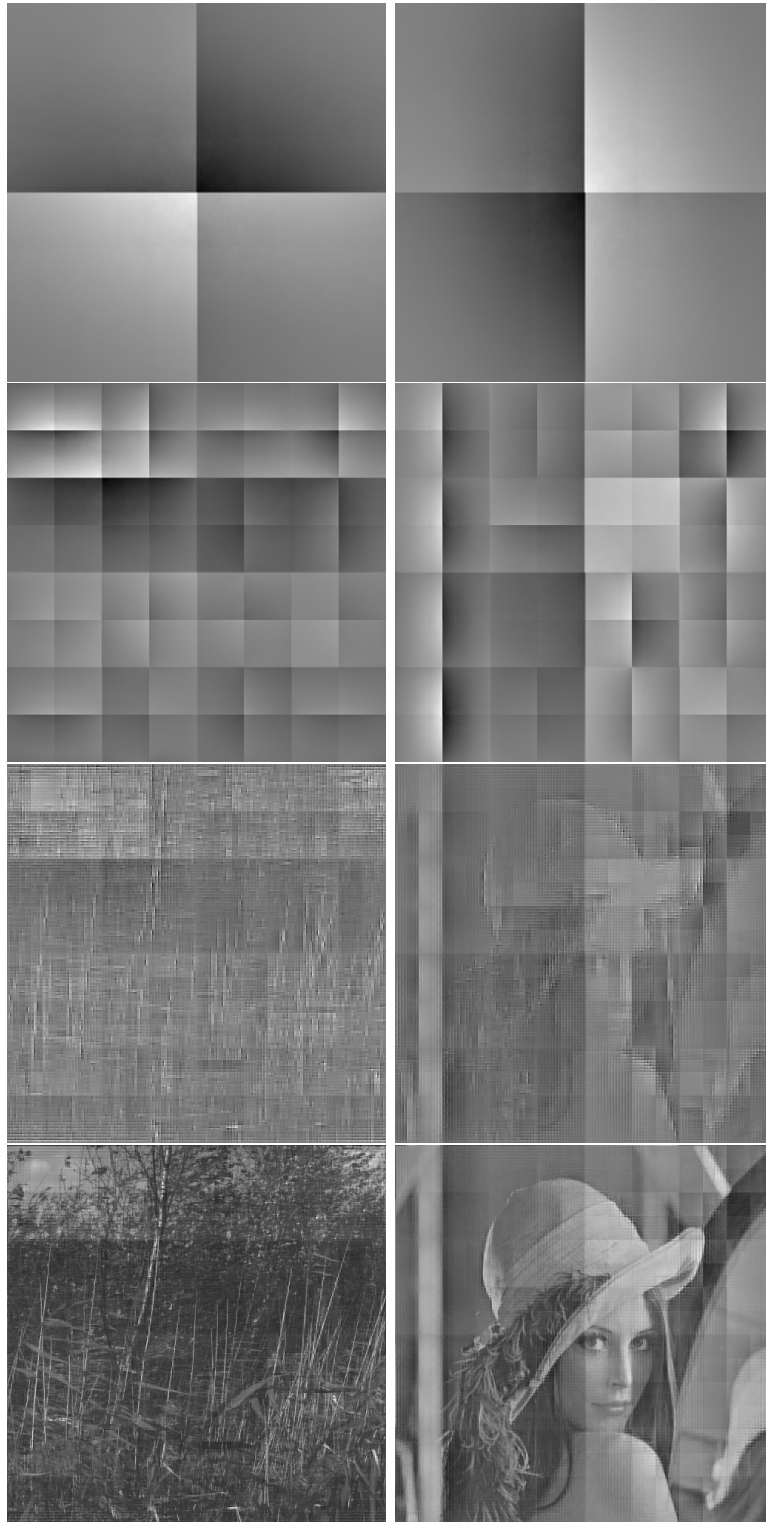


Figura 2.16:  $\sum_{rj\vec{k}} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x})$  para  $j = 0$ ,  $j \leq 2$ ,  $j \leq 6$  y  $j \leq 8$  para las imágenes *imk03236.imc* y de Lena con 2 orientaciones

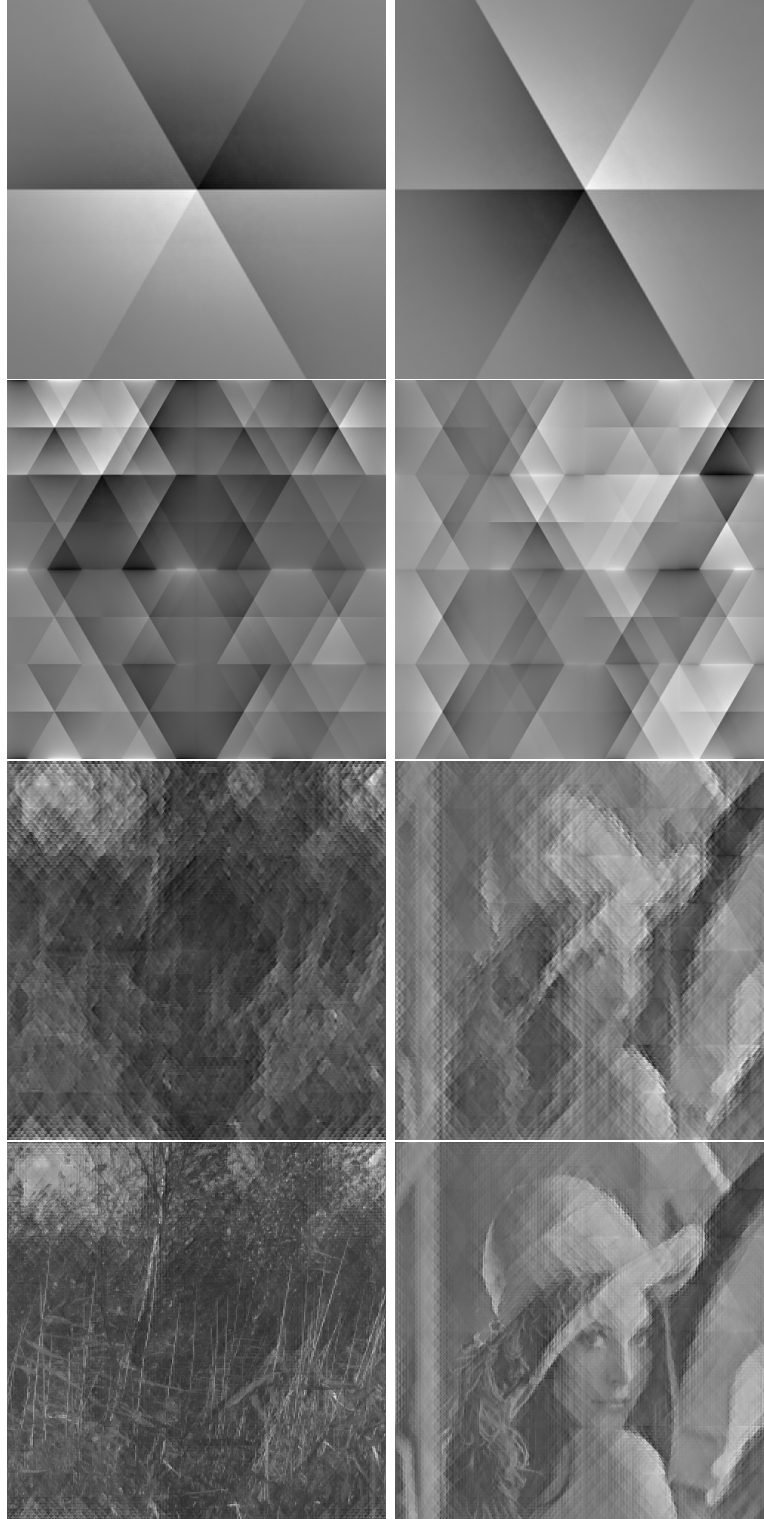


Figura 2.17:  $\sum_{rj\vec{k}} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x})$  para  $j = 0$ ,  $j \leq 2$ ,  $j \leq 6$  y  $j \leq 8$  para las imágenes *imk03236.imc* y de Lena con 3 orientaciones

	1	2	3
<i>imk03236.imc</i>	0.75	1.06	1.45
Lena	0.59	1.17	1.75

Cuadro 2.1: Eficiencias  $\epsilon_n[c]$  para representaciones con  $n = 1, 2$  y  $3$  voces para las imágenes *imk03236.imc* y Lena.

$j$	0	1	2	3	4	5	6	7
$\epsilon_{1j}$	1.000	0.106	0.023	0.006	0.004	0.003	0.003	0.002
$\epsilon_{2j}$	0.001	0.013	0.005	0.001	0.001	0.000	0.000	0.000
$\epsilon_{3j}$	0.085	0.028	0.037	0.025	0.013	0.007	0.003	0.002

Cuadro 2.2: Error medio de la condición de ortogonalidad para la wavelet óptima para las componentes  $r = 0, 1$  y  $2$ .

## 2.9. Conclusiones

Hemos estudiado que las imágenes naturales poseen determinadas regularidades estadísticas. El conocimiento de éstas como los estímulos visuales nos permitirá deducir ciertas propiedades del procesamiento visual. Las simetrías estadísticas que presentan las imágenes naturales son: invariancia de traslación (no hay preferencia de ningún punto en las imágenes), invariancia de escala (los objetos presentes en las imágenes pueden aparecer con un tamaño cualquiera) y multiescala (en las imágenes los objetos que aparecen están jerárquicamente estructurados).

A partir de estas simetrías podemos definir un código eficiente mediante una representación en wavelets que minimiza la redundancia. El análisis en bases de wavelet es el soporte matemático que consigue implementar de manera eficaz las dos primeras simetrías presentes en las imágenes naturales. La wavelet óptima surge de la necesidad de imponer la multiescala al análisis multiresolución. El código que se deriva de esta wavelet óptima descompone las imágenes en niveles independientes de resolución. Además, este código resulta que es esparso sin haber impuesto que lo sea y la wavelet

óptima obtenida actua como un detector de bordes.

Sin embargo, el modelo basado en un análisis en wavelet multiescala no es capaz de resolver el problema de las distintas voces que son necesarias para reconstruir de manera eficiente las imágenes naturales. Aunque una solución *naïf* como el modelo rotacional no es del todo correcta las experiencias muestran que la base real podría ser similar.

## Capítulo 3

# Taxonomía de Imágenes Naturales: Características Visuales

En el capítulo anterior obtuvimos para un colectivo de 4167 imágenes naturales un filtro óptimo -porque extraía las componentes independientes- que funcionaba como detector de bordes horizontales. En este nuevo capítulo aplicaremos lo aprendido a colectivos de imágenes naturales de tamaño más pequeño.

Si para el colectivo de imágenes a escala global obtuvimos un filtro que extraía la característica visual que tenían en común dichas imágenes, con los distintos subcolectivos de subimágenes esperamos obtener diferentes características visuales *locales* que gobernarán la estadística de dichos colectivos. Veremos que los patrones que encontramos en estas subimágenes son bordes, barras, dobles barras, etc, a distintas orientaciones.

Deteminar los colectivos de imágenes naturales que tengan una característica visual que predomine es en sí mismo un problema complejo. Para poder extraer las características visuales que tienen dominancia en un determinado colectivo optamos por una jerarquización de las distintas imagenes de éste. El estudio taxónomico agrupará las imágenes en diferentes clases según la característica visual que más predomine en éstas.

A partir de las distintas características visuales obtenidas de la clasificación de las subimágenes mostraremos dos métodos de reconstrucción de imágenes (a escala global).

Este capítulo presenta los diferentes filtros que aparecen asociados a las distintas

clases (de características visuales) utilizando diferentes métodos que están basados en la fórmula de obtención del filtro óptimo deducida en el capítulo anterior y estudia cómo pueden servir éstos como base de representación de imágenes naturales.

### 3.1. Predominancia de patrones o características visuales

Consideremos la expansión en wavelets del valor medio del contraste de un colectivo de imágenes naturales,

$$\langle c(\vec{x}) \rangle = \sum_{r=0}^{n-1} \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathcal{Z}_{2^j})^2} \langle \alpha_{rj\vec{k}} \rangle \Psi_{rj\vec{k}}(\vec{x}) , \quad (3.1.1)$$

el índice  $r$  que aparece se necesitaba puesto que estamos trabajando con imágenes naturales, y, éstas, como vimos en el capítulo anterior, necesitan varias voces para su completa representación.

Esta búsqueda de las diferentes voces necesarias para la reconstrucción de las imágenes naturales va a ir asociada a la idea de que hay colectivos de imágenes que tienen asociado una determinada característica visual común. Así, para el colectivo de las 4167 imágenes vimos que había una característica visual o patrón común: la existencia de más luminosidad en la parte superior de las imágenes que en la inferior -que no es más que un borde horizontal: un gradiente en el contraste-. A partir de la extracción de dicha característica visual calculamos el filtro óptimo asociado a ésta. Pudimos extraer este filtro porque dicha característica visual se vio que era dominante sobre el resto de las múltiples características visuales presentes en una imagen -véase bordes, barras, dobles barras... en diferentes orientaciones-. Es fácil verificar este argumento viendo el contraste medio del colectivo de las 4167 imágenes en la figura 2.8.

Motivados por los resultados que mostraban que los filtros tipo borde podían ser las componentes independientes de las imágenes naturales [11] y la necesidad de buscar nuevas voces nos llevaron a buscar en diferentes colectivos de imágenes estructuras simples que, mediante nuestra fórmula de obtención del filtro óptimo (que asocia a



cada característica visual una wavelet), pudieran ser los distintos elementos de la base de wavelet.

A partir del argumento de que existen diferentes características visuales en las imágenes naturales, supondremos que el procesamiento visual ha ido adquiriendo éstas, por medio de células simples cuyos campos receptivos serán selectivos a las diferentes características visuales, para una posterior mejor codificación de los estímulos visuales. La relación que encontramos entre los filtros asociados a las características visuales y los campos receptivos de estas células será abordada en el capítulo 5.

En una determinada imagen podemos encontrar que existe una característica visual que a una determinada escala domina sobre las demás, pero que a otra escala esa dominancia no se encuentra. Es decir, dado que estamos haciendo una completa asociación entre componente de wavelet y característica visual, para esa imagen y a esa escala la contribución de la proyección de una wavelet es dominante sobre el resto o, si se quiere, las contribuciones de las otras proyecciones son despreciables. Pensemos en un bosque: a una escala en la que observamos todo el conjunto de árboles, las estructuras de barras y bordes verticales predominarán sobre otras características visuales. Pero si nos acercamos, y vamos a una escala en la que se observe en detalle una hoja de uno de los árboles, entonces las estructuras dominantes ya no serán las anteriores; ahora podrán serlo bordes con una determinada orientación -por el contorno de la hoja-, quizás incluso no se encuentre ninguna característica visual que predomine sobre el resto.

Para poder extraer las diferentes características visuales que existen en un determinado colectivo de imágenes, hacemos uso del promedio del contraste de las imágenes sobre el colectivo y la invariancia de escala. Si a una determinada escala existe alguna característica visual que predomine sobre las demás, entonces por invariancia de escala esa característica visual debe predominar a todas las escalas. Al tratar con el colectivo en vez de con el individuo, podemos hacer uso de la invariancia de escala, y así poder extrapolar a cualquier escala la frecuencia relativa de una determinada característica visual, y en una determinada escala en el colectivo. De tal manera que las características visuales que tengan una frecuencia mayor que otras a una escala mantendrán esa

predominancia a todas las escalas.

Elijamos, entonces, un colectivo de imágenes donde exista una característica visual que sea dominante sobre el resto, lo que implica que hay una voz  $s$  de la base de wavelet que predomina. La ecuación 3.1.1 bajo esta elección de colectivo se podrá aproximar,

$$\langle c(\vec{x}) \rangle_s \sim \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathcal{Z}_{2^j})^2} \langle \alpha_{sj\vec{k}} \rangle \Psi_{sj\vec{k}}(\vec{x}) , \quad (3.1.2)$$

donde el  $\langle \rangle_s$  del primer término denota que ha sido elegido un colectivo de imágenes, las cuales en promedio predomina la wavelet  $s$ . Precisamente por ello, identificamos que ese promedio ha extraído la que hemos venido llamando, pero no definido, *característica visual*.

Si hacemos uso de la multiescala para esta determinada voz, e imponemos que se cumpla punto a punto,

$$\eta_{sj\vec{k}} = \frac{\alpha_{sj\vec{k}}}{\alpha_{sj-1[\frac{\vec{k}}{2}]}} . \quad (3.1.3)$$

podremos llegar a la ecuación de la wavelet óptima para esta determinada componente (haciendo uso de las aproximaciones pertinentes ya conocidas)

$$\Psi_s(\vec{x}) \sim \langle c(\vec{x}) \rangle_s - \langle |\eta_s| \rangle \sum_{\vec{k} \in (\mathcal{Z})^2} \langle c(2\vec{x} - \vec{k}) \rangle_s , \quad (3.1.4)$$

esta relación cabe interpretarla como una aproximación con un primer término dominante -el relacionado con la componente  $s$ - y con los asociados a las otras componentes despreciables. Del capítulo anterior sabemos que la normalización de la wavelet nos impondrá el factor global y que  $\langle |\eta_s| \rangle = \frac{1}{2}$ .

Para deducir este pequeño desarrollo hemos partido de dos hipótesis. Primera, damos por hecho la existencia de una determinada característica visual  $s$  que previamente seleccionamos. Segunda, somos capaces de encontrar un colectivo de imágenes naturales en los que predomine esa característica visual  $s$  para toda escala y posición. Sin embargo, cuando se dedujo la *característica horizonte* no se seleccionó la característica visual previa ni el colectivo y, aún y todo, pudimos extraerla -bien es verdad que ese no era el objetivo-. Sabemos que fue posible porque todas las imágenes del colectivo tenían

una característica visual en común, cuya wavelet asociada la interpretamos como una mezcla de diferentes voces pero que también pudimos interpretarla como una wavelet asociada a una determinada característica (bajo un cambio apropiado de base). El no haber buscado una característica visual seleccionada previamente nos permitió tener una relación como 3.1.4 pero que no es una aproximación sino una verdadera igualdad.

Concluimos que no es necesario la primera de las hipótesis; sin disponer de una base previa podremos llegar a un desarrollo en el que la wavelet mezcla la asociemos con el filtro de una determinada característica visual, de otra forma sería necesario hacer un cambio de base adecuado.

## 3.2. Selección de colectivos y clasificación de las imágenes

Establecer qué colectivos de imágenes naturales tienen una característica visual que predomine sobre el resto es un problema complicado. Se podría optar por construir conjuntos de imágenes con una determinada característica visual, escogiendo cada imagen según viésemos -en sentido literal- en ella si tiene la característica visual elegida. A nivel de codificación sensorial -que abordaremos en un posterior capítulo con la deducción de los campos receptivos de células simples- esta forma de actuar tiene tres grandes inconvenientes.

1. Es una elección supervisada totalmente dependiente del criterio subjetivo de la persona electora de imágenes. Esto va en contra de la hipótesis de que el sistema visual, a lo largo de un aprendizaje, se ha basado en los diferentes estímulos visuales que se encuentra en su entorno para construir un código eficiente que codifique dichos estímulos.
2. Requiere saber a priori qué características visuales son las que desempeñan un papel relevante en la codificación sensorial del proceso visual. Esta cuestión es parte del problema, por tanto hará falta conocer por otros métodos esas características visuales.

3. Por medio de este método no podemos saber cuál es la frecuencia de cada característica visual sobre el cómputo total, es decir, no podemos hacer una clasificación cuantitativa según la abundancia relativa de cada característica visual. Dado que el sistema visual se adapta a la estadística de su entorno, la abundancia relativa de cada característica visual en un colectivo estará estrechamente relacionada con el número de células sensoriales cuyos campos receptivos sean sensibles a dicha característica visual.

Para poder extraer las características visuales que tienen dominancia en un colectivo nosotros optamos por una jerarquización de las distintas imágenes del colectivo.

Nosotros optamos por un método bastante simple: definimos una métrica dentro de nuestro espacio de imágenes y calculamos con esa métrica las distancias entre las distintas imágenes del nuestro colectivo. Una clase vendrá definida por todas las imágenes del colectivo que disten una distancia menor que un cierto umbral de un centro. Utilizamos una relación de equivalencia, por lo tanto, las clases son excluyentes. Las clases serán ordenadas según el número de imágenes clasificadas. Al clasificar las imágenes utilizando la distancia en cada clase existirán dos orientaciones; así, dos imágenes pueden tener orientaciones opuestas aunque pertenezcan a la misma clase. Por ejemplo, si obtenemos una clase que venga definida como borde horizontal, entonces habrá imágenes dentro de esta clase que tendrán más luz en la parte superior que en la inferior, mientras que habrá otras imágenes que les ocurra lo contrario: será en la parte inferior donde haya más luz.

### 3.2.1. Algoritmo de clasificación de imágenes

Se busca una clasificación intrínseca, que separe patrones desconocidos al agrupar imágenes en función de una medida de similitud mutua y que no dependa de un conocimiento a priori de la base de wavelets: precisamente, es en última instancia lo que buscamos. Lo que haremos será construir colectivos de imágenes de acuerdo a la distancia canónica entre dos imágenes, usando una variante del *K-means clustering* [33]. La definición de distancia debe de ser invariante bajo cambios en la iluminación global: se

busca similitud de formas, independiente de su amplitud, porque las wavelets  $\Psi_r$  están por definición normalizadas,  $\langle \Psi_r | \Psi_r \rangle = 1 \forall r$  ( $\langle \Psi_r | \Psi_r \rangle \equiv \int d\vec{x} \Psi_r^2(\vec{x}) = 1$ ). Además, el producto escalar no debería variar cuando se varía el nivel de referencia debido a que las wavelets tienen media nula [27],  $\langle \Psi_r \rangle = 0 \forall r$  ( $\langle \Psi_r \rangle \equiv \int d\vec{x} \Psi_r(\vec{x}) = 0$ ). En nuestro paradigma implementaremos estas simetrías dividiendo cada imagen por su sigma tras sustraerles la media.

Nuestro colectivo a clasificar será el conjunto de todas las subimágenes disjuntas ( $c_i$ ) de una imagen-colectivo ( $I_c$ ) a una escala global dada. Por invariancia de escala podemos, en teoría, escoger el tamaño de estudio que queramos. En la práctica, usaremos pequeñas subimágenes ( $16 \times 16$ ) que son trozos de imágenes ( $1536 \times 1024$ ), por llegar a un compromiso entre la estadística -necesitamos suficientes elementos: 6144- y la estructura de dichos elementos.

Para medir el grado de proximidad de dos subimágenes  $c_i$  y  $c_{i'}$  calcularemos el coeficiente de correlación mutua  $\rho_{ii'}$  que, por definición, es

$$\rho_{ii'} \equiv \frac{\langle c_i | c_{i'} \rangle - \langle c_i \rangle \langle c_{i'} \rangle}{\sigma_i \sigma_{i'}}, \quad (3.2.1)$$

siendo

$$\sigma_i^2 \equiv \langle c_i | c_i \rangle - \langle c_i \rangle^2. \quad (3.2.2)$$

Debemos interpretar en este contexto a  $\langle c_i \rangle$  como la norma de la subimagen  $c_i$  ( $\langle c_i \rangle \equiv \int d\vec{x} c_i(\vec{x})$ ), igual que hicimos antes con  $\Psi_r$ , y no como el promedio del contraste de los diferentes  $c_i$  del colectivo ( $\sum_i c_i$ ). Aunque, por otra parte, sabemos que si los  $c_i$  son *suficientemente* extensos y el número de subimágenes, representativo, por ergodicidad las dos definiciones coinciden.

La correlación  $\rho_{ii'}$  es un coeficiente de solape entre las imágenes (o subimágenes en nuestro caso)  $c_i$  y  $c_{i'}$ , que es invariante bajo cualquier transformación lineal de los niveles de gris debido a su propia definición, como necesitábamos. Este coeficiente puede ser interpretado como una distancia en el espacio de las subimágenes: asumiendo que  $c_i$ ,  $c_{i'}$  están normalizadas (*i.e.*,  $\langle c_i \rangle = \langle c_{i'} \rangle = 0$ ,  $\langle c_i | c_i \rangle = \langle c_{i'} | c_{i'} \rangle = 1$ ), la distancia euclídea entre  $c_i$  y  $c_{i'}$  verifica  $d(c_i, c_{i'}) \equiv \sqrt{\langle (c_i - c_{i'}) | (c_i - c_{i'}) \rangle} = \sqrt{2(1 - \rho_{ii'})}$ . Así,

imágenes muy *próximas* ( $d(c_i, c_{i'}) \rightarrow 0$ ) tendrán un coeficiente de correlación cercano a uno ( $\rho_{ii'} \rightarrow 1$ ).

Una vez definida la correlación, el paso siguiente para construir las clases será fijar un umbral de solape de referencia  $0 < Q \leq 1$  que consideremos que es suficiente para considerar a dos subimágenes razonablemente similares, esto es, que pertenecen a la misma clase. Este umbral lo podemos asociar también a dos formas geométricas distintas: bien con el radio de la bola con centro en cada subimagen de la cual queremos estudiar su clasterización, y también con el ángulo que forman las distintas subimágenes de una clase, dentro de este espacio, (este ángulo no puede superar ese umbral). En nuestro estudio lo fijamos en  $Q = 0.3$ . El umbral de solape escogido (lo que llamamos  $Q$ ) puede parecer bajo, esto es, somos poco exigentes a la hora de fijar la frontera que delimita qué subimágenes pertenecen a una determinada clase, y en ésta tendrían cabida muchos elementos que, en principio, serían bastante diferentes. Sin embargo, debido a que estamos trabajando en un espacio de dimensión 256 (imágenes de  $16 \times 16$ ) se puede ver que no es tan bajo. En la figura 3.1 se muestra como a medida que aumenta la dimensión del espacio la fracción de vectores contenidos sobre el total ( $\Lambda$ ) en función del solape escogido ( $Q$ ) disminuye. Podemos explicarlo de otro modo: para imágenes independientes de  $16 \times 16$  píxeles el solape promedio es 0 y la desviación estándar respecto a este valor medio es de forma aproximada  $1/\sqrt{256} = 1/16 = 0.0625$ . Por tanto, estamos escogiendo un solape que se desvía al menos 5 desviaciones estándar respecto al modelo independiente; véase además que con la estadística que tenemos  $5\sigma$ , aunque efectivamente las imágenes no siguen una ley normal, supone recortar aún y todo considerablemente el muestreo disponible.

En la figura 3.3 mostramos los distintos mapas con las subimágenes de tamaño  $16 \times 16$  para cada una de las clases (tomamos las cinco más representadas) que han sido seleccionadas de la imagen-colectivo *imk00017.imc* que podemos ver en la figura 3.2. Para ilustrar lo discutido en el parrafo anterior hemos clasificado los elementos utilizando dos diferentes solapes,  $Q = 0.3$  y  $Q = 0.7$ .

Una vez que hemos fijado el nivel de solape, definimos el  $Q$ -entorno de una subimagen  $c_i$ , que denotamos por  $U_Q(c_i)$ , como el conjunto de todas aquellas imágenes que

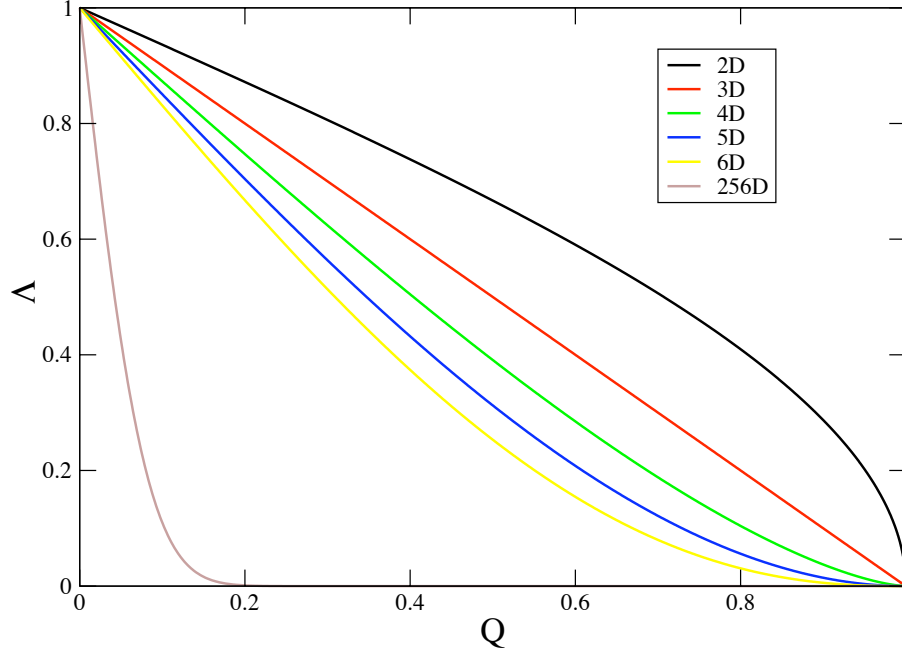


Figura 3.1: Fracción de vectores ( $\Lambda$ ) en función del solape ( $Q$ ) para espacios vectoriales de distintas dimensiones (2, 3, 4, 5, 6 y 256).

tienen un solape mayor o igual que  $Q$  con  $c_i$  o con  $-c_i$ , es decir:

$$U_Q(c_i) = \{c_{i'} : |\rho_{ii'}| \geq Q\}. \quad (3.2.3)$$

Es importante hacer notar que se ha tomado el valor absoluto del coeficiente  $\rho_{ii'}$  en la definición de  $Q$ -entorno para incluir no sólo las imágenes que están próximas a  $c_i$  sino también aquellas que lo están de la imagen  $-c_i$ . De esta manera, aumentamos el número de subimágenes de las clases al suprimir un grado de libertad trivial que no tiene relevancia visualmente (un cambio en el signo implica una inversión en el gradiente de contraste, que puede ocurrir igualmente en un sentido o en el contrario). En términos de distancias euclídeas se puede decir de forma equivalente que  $U_Q(c_i)$  consiste de aquellas imágenes tales que su distancia euclídea a  $c_i$  o a  $-c_i$  es menor o igual a  $\sqrt{2(1-Q)}$  (o sea, es la unión de la bola de centro  $c_i$  y radio  $\sqrt{2(1-Q)}$  con



Figura 3.2: Imagen *imk00017.imc* de la base de datos de Hans van Hateren (ver [30] para detalles sobre la imagen).

la bola de centro  $-c_i$  y mismo radio). De la misma forma, podemos pensar en  $U_Q(c_i)$  como el conjunto de todas las imagenes tales que el ángulo que forman con  $c_i$  o con  $-c_i$  es menor que  $\cos^{-1}(Q)$ .

El número de subimágenes incluídas en el  $Q$ -entorno de una subimagen  $c_i$  es una medida del grado de vecindad (*clustering*) de dicha subimagen. Entonces, definimos el grado de vecindad de  $c_i$ ,  $V_Q(c_i)$ , de la forma siguiente

$$V_Q(c_i) = \#\{i' : |\rho_{ii'}| \geq Q\}. \quad (3.2.4)$$

Diremos que una imagen  $c_{i_c}$  es central de la clase más representada si su grado de vecindad es superior al de cualquier otra imagen,

$$V_Q(c_{i_c}) > V_Q(c_i) \forall i \neq i_c. \quad (3.2.5)$$

Para construir las distintas clases se procede de forma iterativa. Primero, obtenemos la imagen central de la primera clase más representada, que denotamos por  $c_{i_0}$ . La primera clase  $\mathcal{C}_0$  se identifica con con el  $Q$ -entorno de  $c_{i_0}$ :  $\mathcal{C}_0 \equiv U_Q(c_{i_0})$ .



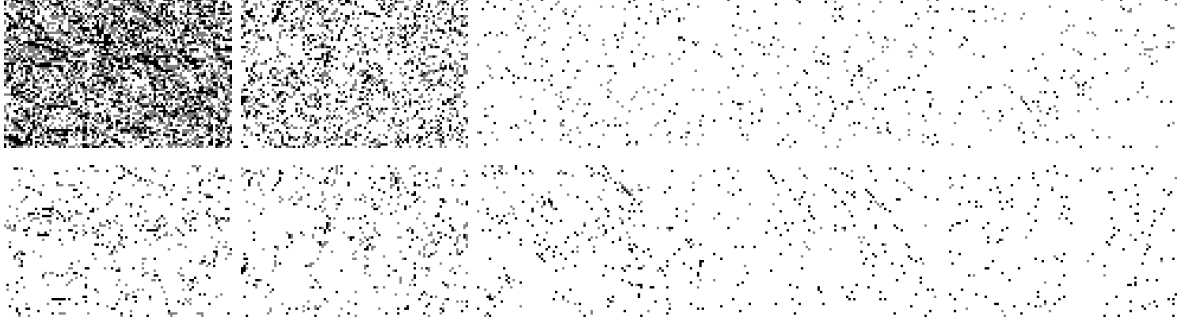


Figura 3.3: Mapas con los elementos  $16 \times 16$  seleccionados para cada una de las cinco clases más representadas de la imagen *imk00017.imc* con  $Q = 0.3$  (**arriba**) y con  $Q = 0.7$  (**abajo**). Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados.

### 3.2.2. Extracción de las características visuales

Una vez que hemos construido la primera clase,  $\mathcal{C}_0$ , podemos calcular la característica visual asociada como la suma signada de todas las imágenes de  $U_Q(c_{i_0})$  (se usan las medias signadas para reducir los efectos de fluctuación estadística; esto será discutido en detalle en el siguiente capítulo); i.e.,

$$\langle c \rangle]_0 = \sum_{c_i \in \mathcal{C}_0} \text{sign}(\rho_{i_0 i}) c_i. \quad (3.2.6)$$

Tenemos que incluir en el sumatorio tanto las imágenes próximas a  $c_{i_0}$  como las próximas a  $-c_{i_0}$ .

Para obtener la segunda clase más representada,  $\mathcal{C}_1$ , excluirémos todas las subimágenes de  $U_Q(c_{i_0})$  y buscaremos de nuevo la subimagen central, denotada por  $c_{i_1}$ . La segunda clase se construye inmediatamente a partir de ella:  $\mathcal{C}_1 \equiv U_Q(c_{i_1}) - U_Q(c_{i_0}) = U_Q(c_{i_1}) - \mathcal{C}_0$ , y la característica visual asociada es:

$$\langle c \rangle]_1 = \sum_{c_i \in \mathcal{C}_1} \text{sign}(\rho_{i_1 i}) c_i \quad (3.2.7)$$

y así iteramos para obtener las clases restantes, de acuerdo con la siguiente expresión para la clase  $s + 1$ -ésima:

$$\mathcal{C}_s \equiv U_Q(c_{i_s}) - \cup_{r=0,\dots,s-1} \mathcal{C}_r \quad (3.2.8)$$

y la  $s + 1$ -ésima característica visual:

$$\langle c \rangle_s = \sum_{c_i \in \mathcal{C}_s} \text{sign}(\rho_{i_s i}) c_i. \quad (3.2.9)$$

### 3.2.3. Filtro óptimo asociado a la característica visual

Después de haber clasificado las imágenes (o las partes de ellas) dentro del colectivo y asociar a cada clase una característica visual supondremos que las imágenes de dicho colectivo también cumplen el proceso multiplicativo. Estaremos en una situación similar a la que nos encontramos cuando hayamos el filtro óptimo para las 4167 imágenes, aunque ahora hayamos elegido de manera *artificial* los distintos colectivos. Podemos, por tanto, calcular el filtro óptimo asociado a cada característica visual mediante la ecuación 3.1.4.

En el siguiente capítulo veremos una deducción formal que justifique por qué nuestro algoritmo de clasificación es capaz de separar las distintas voces de una base de wavelet, por qué asociamos las características visuales con las distintas componentes de dicha base y *last, but not least* por qué se sigue cumpliendo el proceso multiplicativo en cada colectivo de imágenes.

## 3.3. Implementación del algoritmo de clasificación

En las figuras 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11 y 3.12 se muestran diferentes paradigmas -obviamente, se pretende que sean representativos- de cómo se clasifican las imágenes dentro de un colectivo si se usa el método descrito en la sección anterior. Hemos usado como colectivo partes de imágenes pertenecientes todas a la misma imagen. Las clases están representadas por un código de tres colores: el blanco significa

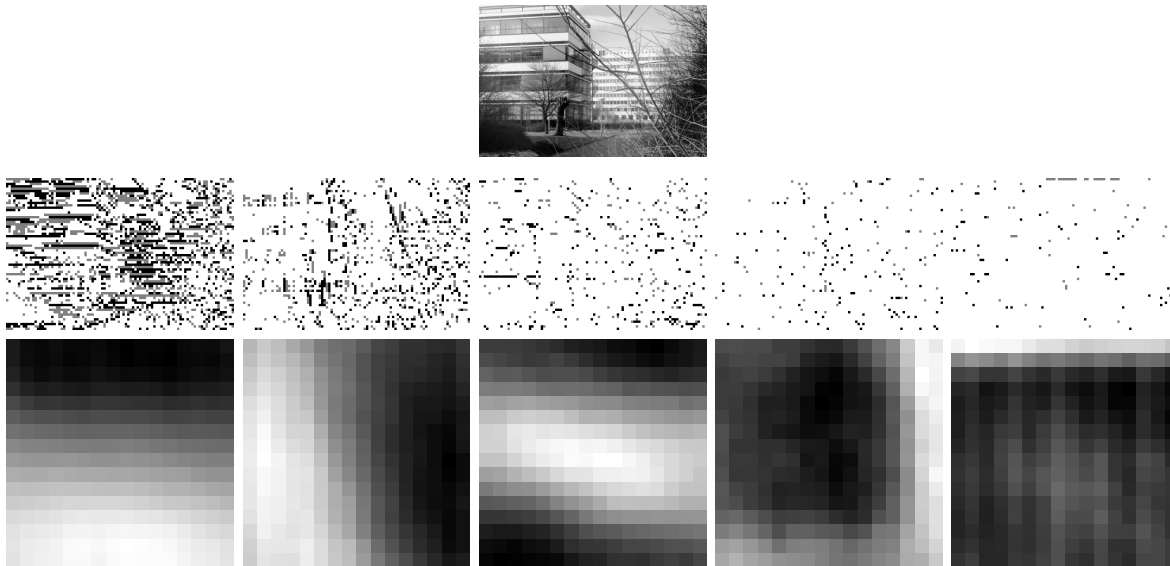


Figura 3.4: **Arriba:** Imagen-colectivo *imk00005.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

las subimágenes que no pertenecen a la clase, con el negro representamos las imágenes que perteneciendo a la clase tienen todas la misma orientación, convencionalmente positiva, y el gris lo reservamos para todas aquellas imágenes pertenecientes a la clase que tienen orientación opuesta a la del caso anterior, negativa. También se muestra las características visuales de cada clase, éstas han sido obtenidas promediando los elementos de cada clase, y teniendo en cuenta el signo de cada imagen dentro de la clase.

Podemos extraer una serie de observaciones de las figuras presentadas en el párrafo anterior. Primero, de las cinco clases más dominantes seleccionadas, las tres primeras suelen tener asociadas características visuales con suficiente nitidez; tienen una estadística -el número de subimágenes en la clase es *abundante*- tal que el promedio signado dé algo con coherencia: predomina la característica visual que intentamos ex-

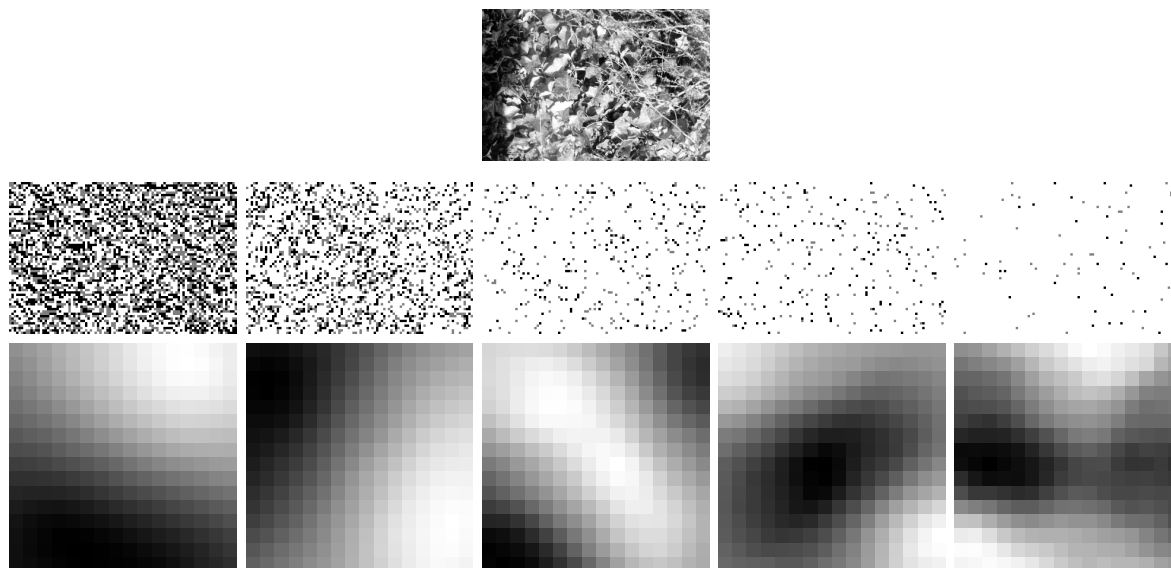


Figura 3.5: **Arriba:** Imagen-colectivo *imk00016.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

traer sobre el ruido producto de las fluctuaciones estadísticas. Evidentemente, estamos siendo subjetivos al hablar de *suficiente nitidez*, pero el hecho de que podamos apreciar simples objetos bien definidos -eso no se puede poner en duda- ya nos da una idea de la robustez de que el algoritmo nos está dando *algo*. Sin embargo, las otras características visuales asociadas a las otras clases, en general, vienen dominada por las fluctuaciones, y no se pueden distinguir patrones con claridad.

Segundo, centrándonos en estas tres primeras características visuales encontramos una amplia gama de patrones simples en distintas orientaciones (bordes: una única transición en el contraste -que pasa de más claro a más oscuro- en una determinada dirección, barras: dos transiciones en el contraste en una determinada dirección, círculos: transición del contraste con simetría radial). El tamaño de subimagen escogido ( $16 \times 16$ ) provoca una limitación en la complejidad de estos patrones, nos es difícil

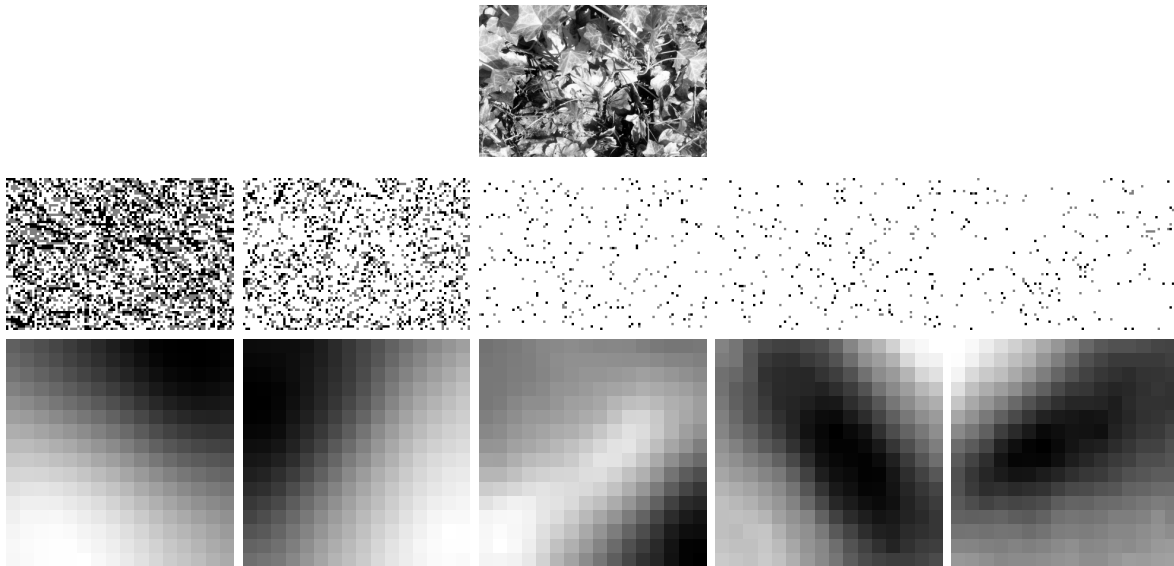


Figura 3.6: **Arriba:** Imagen-colectivo *imk00017.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

encontrar, por ejemplo, dobles barras: estructuras donde exista un doble gradiente en el contraste.

Tercero, las características visuales (que son de escala prefijada) suelen aparecer co-alineadas según su dirección principal (i.e., los bordes y barras horizontales suelen aparecer en grupos alineados horizontalmente, etc). En la figura 3.12 podemos ver cómo con sólo la primera clase somos capaces de reconstruir parcialmente la imagen original.

En la figura 3.24 representamos la clasificación de las cinco primeras características visuales de 500 imágenes seleccionadas: en total, pues, tenemos 2500 características visuales.

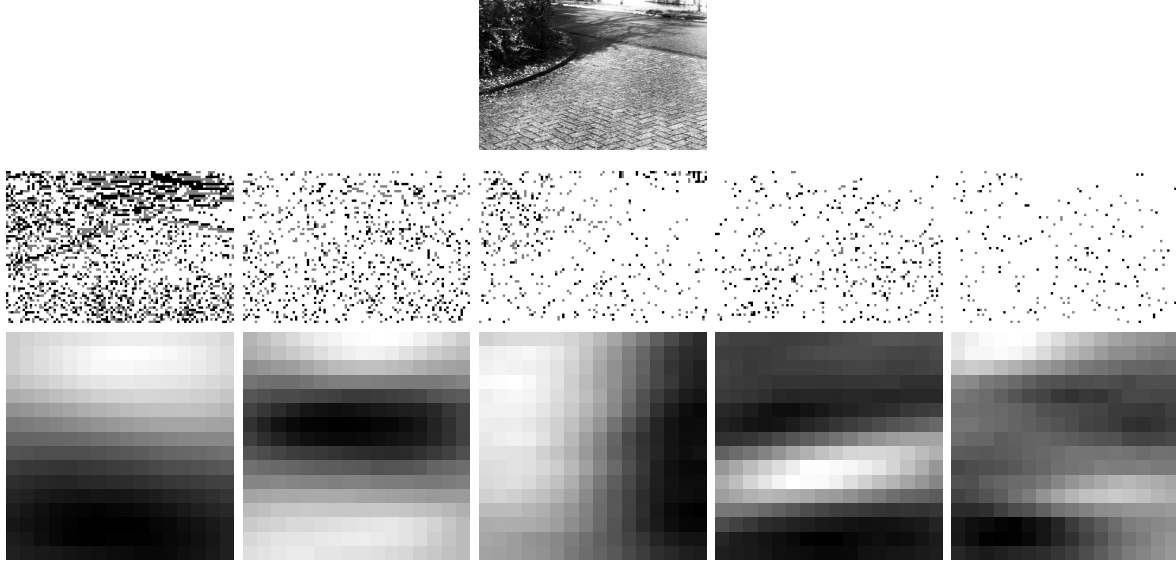


Figura 3.7: **Arriba:** Imagen-colectivo *imk00019.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

### 3.3.1. Aplicación del algoritmo en la representación de imágenes

A modo de simple test, [34], hemos probado qué tipo de representación se consigue cuando se usan separadamente los filtros asociados (derivados a partir de la ecuación 3.1.4) a cada una de las 5 características visuales predominantes (5 imágenes de representación) o cuando se combinan no linealmente: el ganador se lo lleva todo y en cada posición y escala se selecciona uno sólo de los filtros asociados a las 5 características visuales, el de amplitud mayor.

Se ha representado la imagen  $I_c$  que podemos observar en la figura 3.13 utilizando los filtros, figura 3.14, que obtenemos de sus características visuales más predominantes, figura 3.11. Para cada característica visual  $r$  podemos definir su *representación* (de la

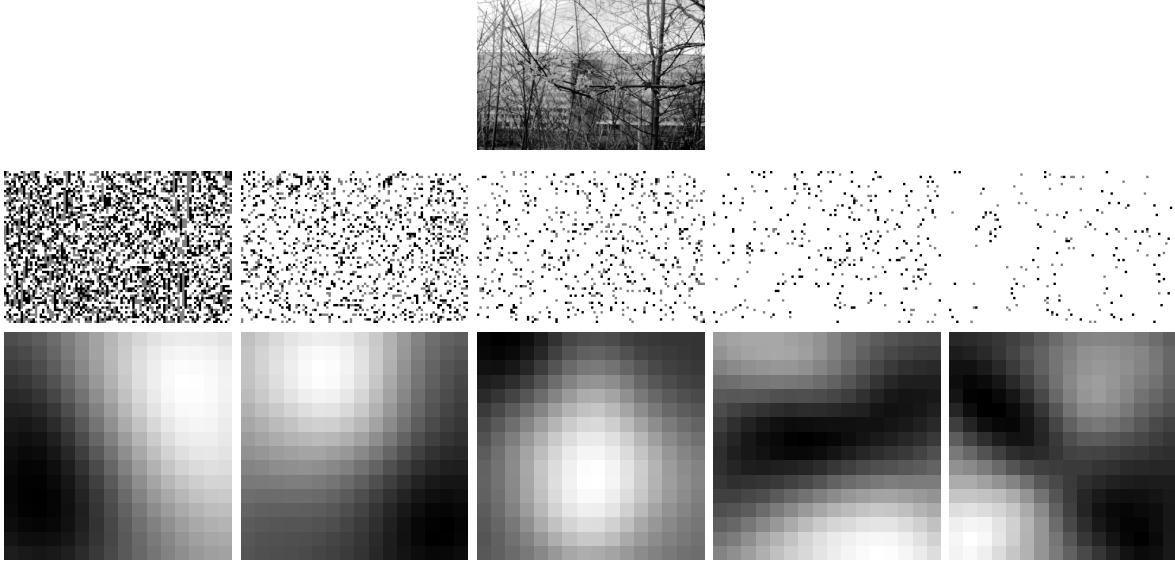


Figura 3.8: **Arriba:** Imagen-colectivo *imk00207.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

imagen anterior) asociada, mediante  $R_r(\vec{x})$ ,

$$R_r(\vec{x}) = \sum_j \sum_{\vec{k}} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x}), \quad (3.3.1)$$

siendo los  $\alpha_{rj\vec{k}}$ , como siempre, las proyecciones  $\langle I_c | \Psi_{rj\vec{k}} \rangle$ . En la figura 3.14 se pueden observar las distintas representaciones.

Podemos construir de igual forma la representación óptima  $R_{opt}(\vec{x})$  (no se debe interpretar ésta en el sentido que le estamos dando al filtro óptimo: aquel que es capaz de extraer las componentes independientes) usando los cinco filtros y seleccionando a cada escala y posición el filtro que mejor ajusta la imagen a esa escala y posición,

$$R_{opt}(\vec{x}) = \sum_j \sum_{\vec{k}} \alpha_{opt(j,\vec{k})j\vec{k}} \Psi_{j\vec{k}}^{opt(j,\vec{k})}(\vec{x}), \quad (3.3.2)$$

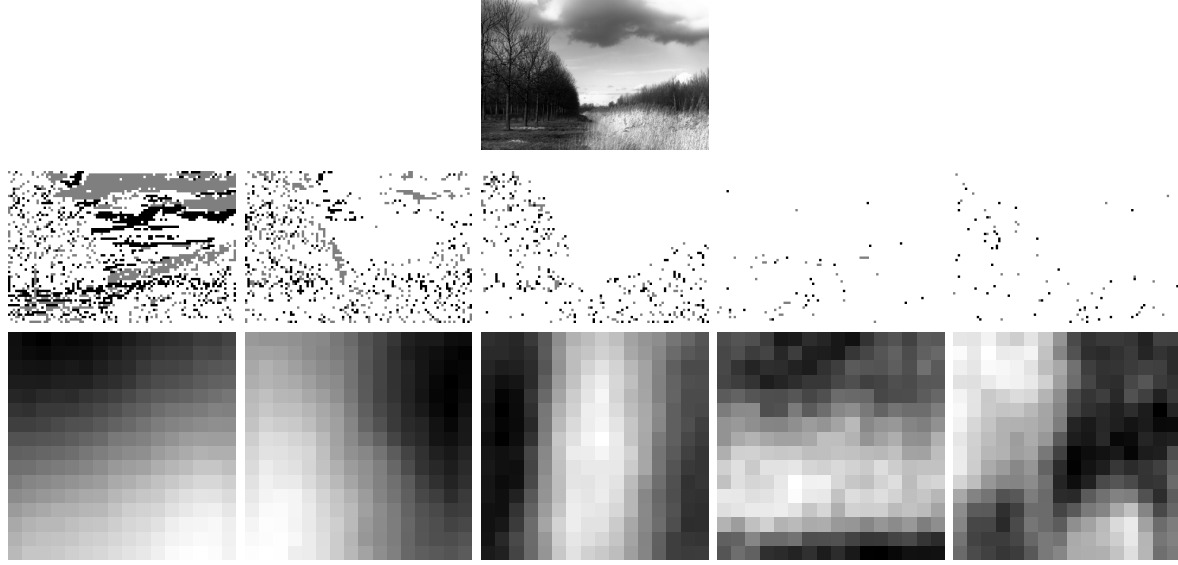


Figura 3.9: **Arriba:** Imagen-colectivo *imk00406.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

donde, para cada  $j$  y  $\vec{k}$ , tomamos como wavelet *mejor adaptada*  $\Psi_{j\vec{k}}^{opt(j,\vec{k})} \equiv \Psi_{sj\vec{k}}$  tal que cumple la condición:  $|\alpha_{sj\vec{k}}| > |\alpha_{rj\vec{k}}| \forall r \neq s$ . En la figura 3.15 se muestra esta representación.

Para evaluar la calidad de estas representaciones, calculamos la imagen de error,

$$\epsilon(\vec{x}) = I_c(\vec{x}) - R_{approx}(\vec{x}), \quad (3.3.3)$$

donde  $I_c(\vec{x})$  es el contraste de la imagen a representar y  $R_{approx}(\vec{x})$  es o uno de los  $R_r$  o  $R_{opt}$ , y así podemos evaluar la calidad de la aproximación usando el PSNR, *peak signal-to-noise ratio*, que se define como

$$PSNR \equiv 10 \cdot \log_{10}\left(\frac{R_D^2}{\sigma_\epsilon^2}\right), \quad (3.3.4)$$

siendo  $R_D$  el rango dinámico de la imagen original y  $\sigma_\epsilon$  la desviación típica de la



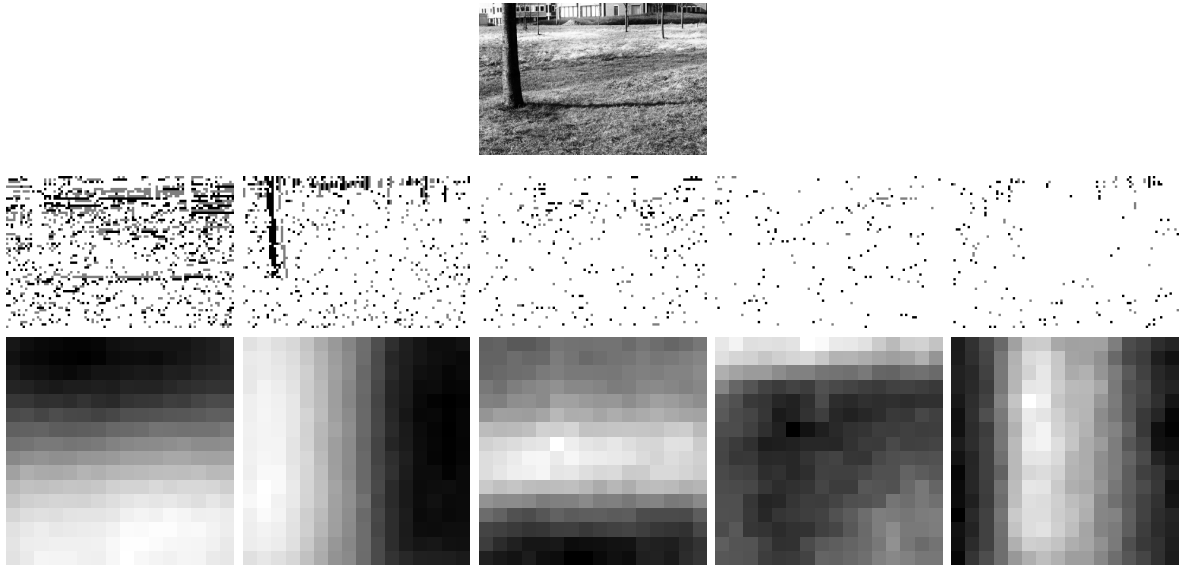


Figura 3.10: **Arriba:** Imagen-colectivo *imk00408.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

imagen de error. Los beneficios de usar bancos de filtros en vez de filtros individuales son evidentes, ya que los bancos implican una reducción sustancial del ruido.

Aunque nuestro propósito no era conseguir un buen método de representación con este sencillo test se ve la necesidad de diversas voces y que el uso de una base sobre-completa podría aligerar el peso del código, aún cuando un código basado en  $\eta$  es esparso, como ya se argumentó en el capítulo anterior.

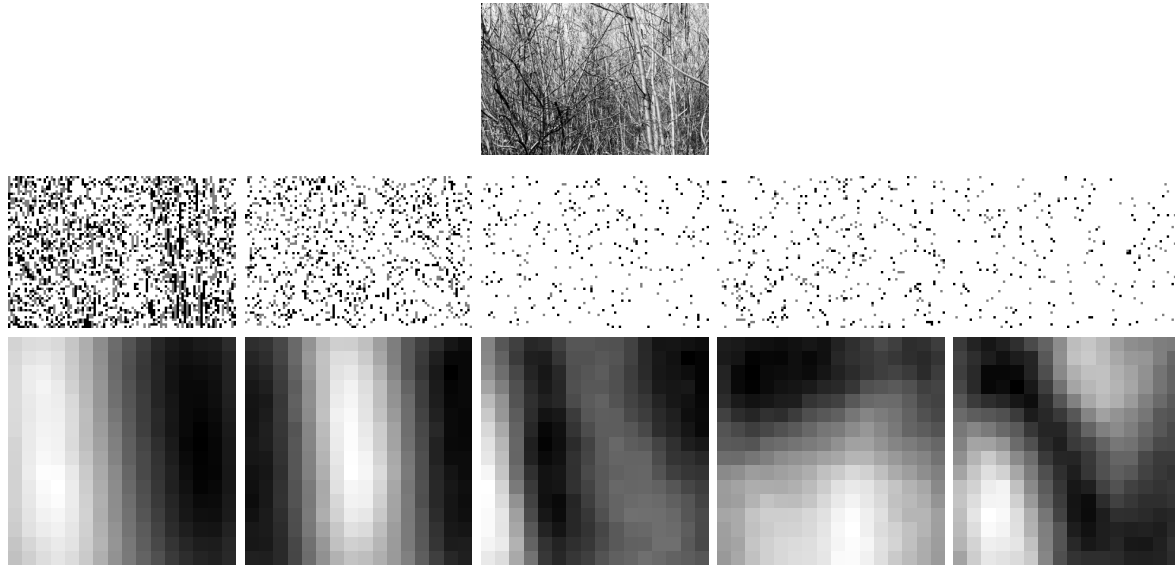


Figura 3.11: **Arriba:** Imagen-colectivo *imk00802.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

### 3.4. Limitación de la fórmula de la wavelet óptima: modelo analítico

Hemos comentado anteriormente los problemas -y sobre todo en la sección del apéndice A.3 se discute con profundidad- que tenemos al evaluar la fórmula de la wavelet óptima, causados por nuestra ignorancia de parte de la señal. La aplicación de la fórmula 3.1.4 (o su dual en el espacio de Fourier), de la misma manera que lo hemos venido haciendo, a las características visuales obtenidas, no nos da unos resultados aceptables por varios motivos:

1. Las imágenes que utilizamos en el colectivo son relativamente pequeñas ( $16 \times 16$  píxeles). Este tamaño fue escogido para asegurar la extracción de características

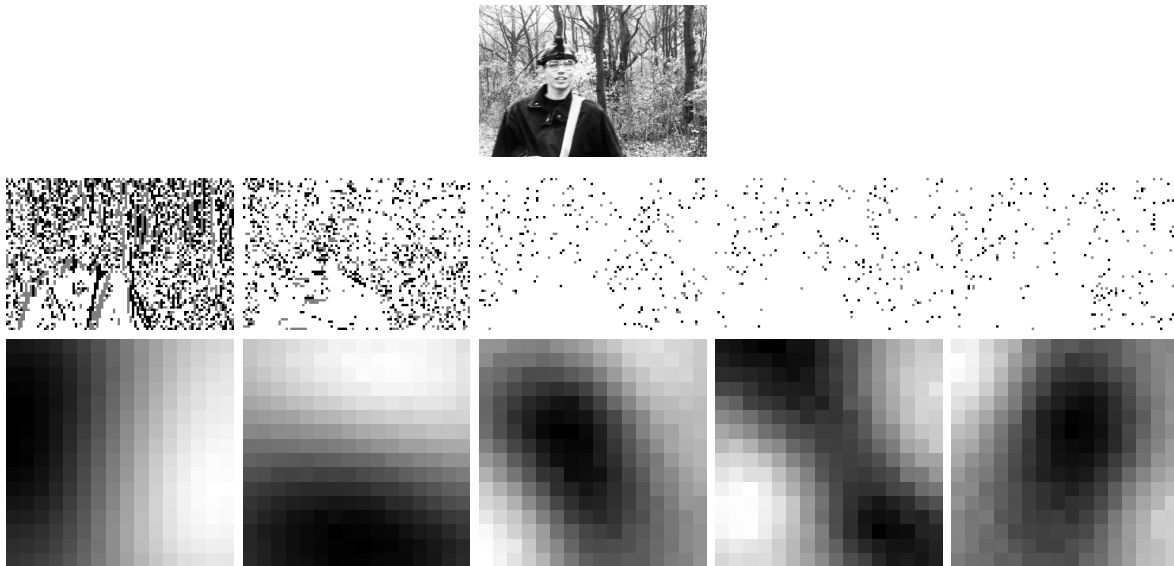


Figura 3.12: **Arriba:** Imagen-colectivo *imk03813.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluídos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central.

visuales simples en las imágenes, pero tiene la desventaja que con tan poca información se hace más evidente la dependencia en el método de extrapolación de la información que nos falta. En la figura 3.14 pudimos observar ejemplos de wavelet obtenidos en el espacio directo a partir de sus correspondientes características visuales.

2. Debido a que en todo momento estamos trabajando con bases diádicas, las wavelets obtenidas a partir de características visuales diagonales no quedan bien representadas. La base diádica fija unos ejes principales -horizontal y vertical-, de tal modo que cuando nos encontramos con una característica visual tipo borde o barra con una orientación que no coincida con dichos ejes, la wavelet resultante será la combinación -con un cierto peso- de los dos ejes: una cruz. En las ca-



Figura 3.13: Imagen *imk00802.imc* de la base de datos de Hans van Hateren (ver [30] para detalles sobre la imagen).

racterísticas visuales cuyo eje no coincida con el definido por la base diádica se acentuará más el problema del punto anterior. En la figura 3.16 mostramos las wavelets obtenidas de aplicar la fórmula en el espacio directo y en el espacio de Fourier a la característica visual que se extrajo de la primera clase de la imagen *imk00017.imc*.

Para poder resolver parte de estos problemas nosotros proponemos como alternativa un modelo analítico. Tomando los resultados condensados en la Figura 3.24, pretendemos derivar un modelo analítico para describir características visuales. Las características visuales que hemos venido mostrando suelen tener una dirección de máxima variación en el contraste: en esencia es lo que define que la característica visual sea horizontal, vertical o, más generalmente, diagonal (para este estudio analítico no consideraremos la característica visuales con simetría circular). Considerar la hipótesis de que la dirección perpendicular a la de máxima variación tiene un decaimiento que parametrizaremos con una gaussiana está justificado por la invariancia traslacional -que se puede observar en los ejemplos mostrados-.

Después de un estudio cualitativo en diferentes imágenes, propusimos una sencilla parametrización para ese conjunto basada en la similitud y simplicidad de los perfiles:

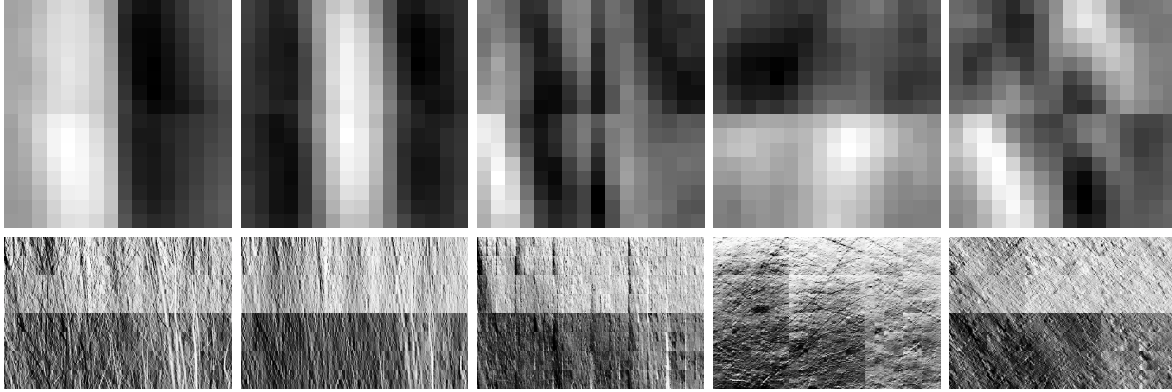


Figura 3.14: Imágenes de las representaciones (**abajo**) de la imagen *imk00802.imc* usando los filtros (**arriba**) derivados de las características visuales de las cinco clases más representativas de esa misma imagen. Los PSNR individuales son 23.08 dB, 21.66 dB, 21.16 dB, 21.58 dB y 21.23 dB, respectivamente.

típicamente tienen una, dos o incluso tres transiciones de contraste (*i.e.*, rápidamente pasan de brillo a oscuridad o lo contrario), y luego decaen asintóticamente. En las figuras 3.17, 3.18, 3.19 y 3.20 podemos ver cortes prototípicos de características visuales. Hay que decir que nos basamos en puros aspectos fenomenológicos y sólo, de momento, de manera cualitativa. Pero la idea de poder parametrizar el conjunto de las características visuales es bastante atractiva *per se*, en la medida que nos permite tratar como conjunto a todas ellas y no perderse en los detalles secundarios de cada característica visual.

### 3.4.1. Modelo sigmoide

En el artículo [35] mostramos que los diferentes cortes -en la dirección de máxima variación del contraste- de las características visuales más simples revelan que la parte central evoca a una sigmoide, que podemos aproximar bien con una tangente hiperbólica. Sea  $x$  la variable de posición a lo largo de la dirección de máxima variación y coloquemos la transición en el origen de  $x$ ; tenemos pues:

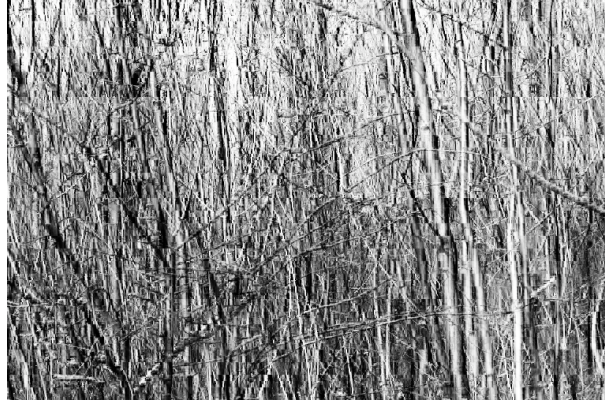


Figura 3.15: Representación usando el banco entero de filtros. El PSNR para esta representación es de 24.27 dB.

$$F^0(x) \approx \tanh(\beta x) \quad |x| \ll 1. \quad (3.4.1)$$

Pero esto sólo aproxima el comportamiento alrededor de la parte central. Para incluir también el decaimiento, proponemos una corrección basada en funciones de la misma familia: tangentes hiperbólicas.

$$f^0(x) = \tanh(\beta x) - \tanh(\beta_\infty x), \quad (3.4.2)$$

donde  $\beta < \beta_\infty$ . Mediante esta primera formulación parametrizamos los cortes de las características visuales con una única transición: los bordes, presentes en nuestra clasificación. Asignando el valor 1 a la longitud total de los cortes, se puede ver que los valores de  $\beta$  toman valores entre 2 y 4. De la misma manera podemos considerar que  $\beta_\infty = 4$ .

Parametrizaremos los cortes de características visuales con una mayor número de transiciones añadiendo estructuras de tipo  $f^0$ . De tal forma, nuestro modelo para dos transiciones -las barras- será

$$f^1(x) = \tanh(\beta(x + x_0)) - \tanh(\beta(x - x_0)), \quad (3.4.3)$$

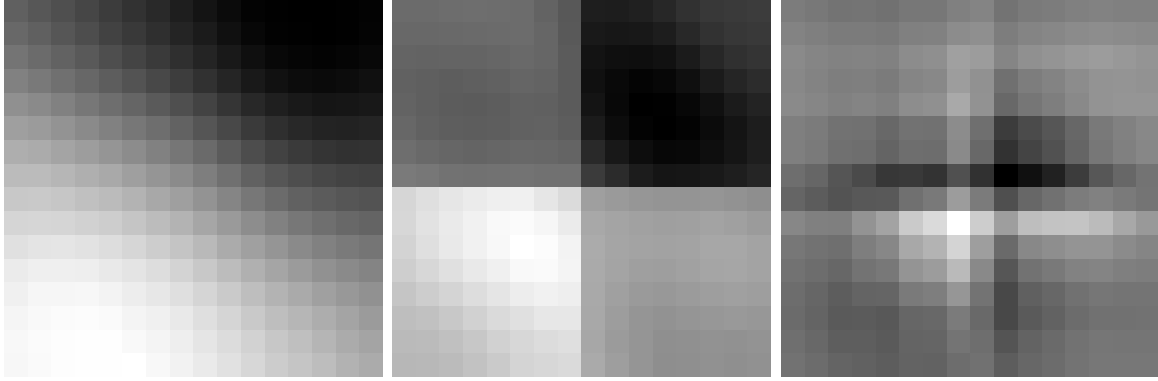


Figura 3.16: **Izquierda:** Característica visual de la primera clase de la imagen *imk00017.imc*. **Centro:** Wavelet obtenida aplicando la fórmula en el espacio de directo. **Derecha:** Wavelet obtenida aplicando la fórmula en espacio de Fourier.

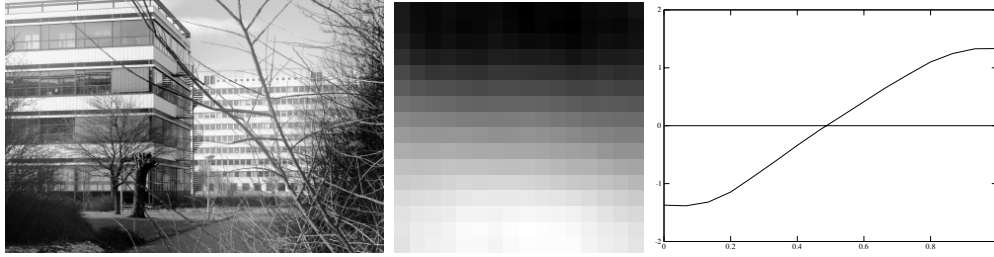


Figura 3.17: Imagen *imk00005.imc* (izquierda), característica visual asociada a la primera clase (centro) y su corte vertical (derecha).

donde, como anteriormente, se asume que la coordenada  $x$  tiene su origen en el centro de la característica visual. Para estos cortes podemos asumir un valor típico de  $x_0$  alrededor de 0.1, mientras que  $\beta$  va de 2 a 4, como antes.

Finalmente, se puede definir una parametrización para tres transiciones:

$$f^2(x) = -\frac{1}{4} \tanh(\beta_\infty(x - x_\infty)) + \tanh(\beta(x - x_0)) - \tanh(\beta(x + x_0)) + \frac{1}{4} \tanh(\beta_\infty(x + x_\infty)), \quad (3.4.4)$$

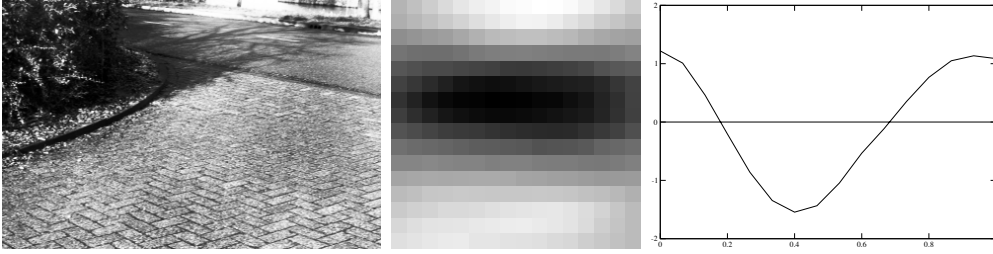


Figura 3.18: Imagen *imk00019.imc* (izquierda), característica visual asociada a la segunda clase (centro) y su corte vertical (derecha).

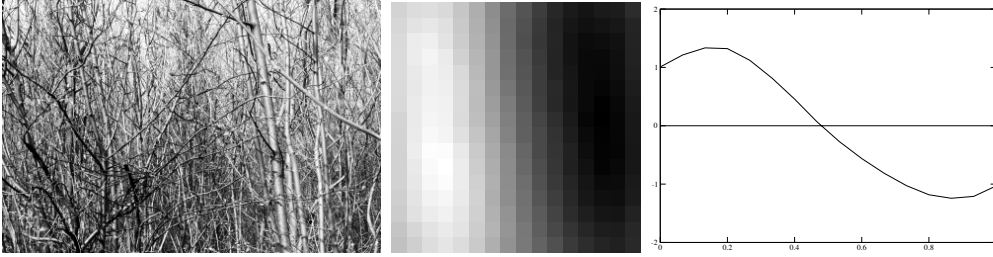


Figura 3.19: Imagen *imk00802.imc* (izquierda), característica visual asociada a la primera clase (centro) y su corte horizontal (derecha).

con  $x_0 < x_\infty$ . Valores típicos son:  $x_0$  de 0.1 a 0.3,  $x_\infty$  de  $x_0$  a 0.4,  $\beta$  de 2 a 4 y  $\beta_\infty$  alrededor de 4. En la Figura 3.21 mostramos las gráficas de  $f^0$ ,  $f^1$  y  $f^2$ .

Para completar la parametrización de las características visuales debemos proporcionar una continuación razonable a lo largo de la dirección perpendicular. Como se argumentó al principio de la sección, escogeremos una gaussiana para extender las características visuales de la siguiente forma:

$$\langle c(x, y) \rangle_r = f^r(x) e^{-\frac{1}{2} \frac{y^2}{\sigma_y^2}}, \quad (3.4.5)$$

donde, como antes, hemos asumido que la coordenada perpendicular está centrada en la característica visual (aunque ahora ya no se trate de un promedio mantenemos la misma notación:  $\langle c(x, y) \rangle_r$ ). Si asignamos un valor convencional unitario a la longitud



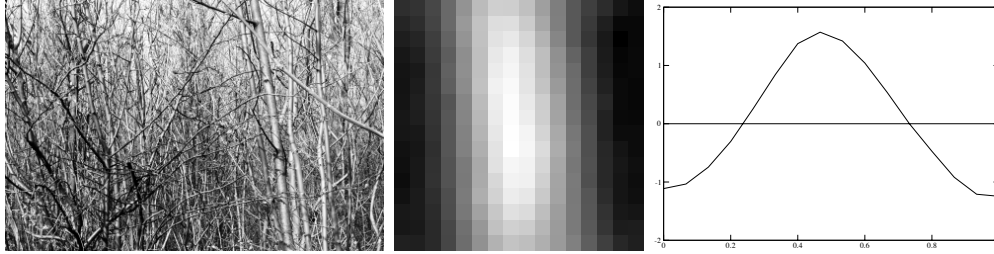


Figura 3.20: Imagen *imk00802.imc* (izquierda), característica visual asociada a la segunda clase (centro) y su corte horizontal (derecha).

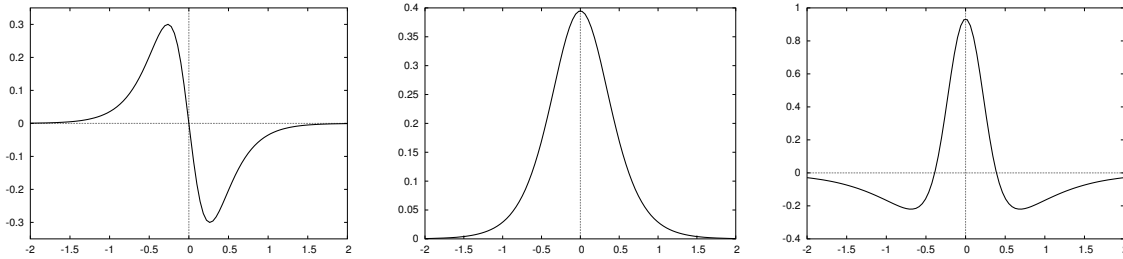


Figura 3.21: Cortes analíticos típicos del modelo.  $f^0$  con  $\beta = 2$ ,  $\beta_\infty = 4$  (**izquierda**),  $f^1$  con  $\beta = 2$ ,  $x_0 = 0.1$  (**centro**) y  $f^2$  con  $\beta = 1$ ,  $x_0 = 0.2$ ,  $\beta_\infty = 4$ ,  $x_\infty = 0.2$  (**derecha**).

de la característica visual en la dirección transversal, un valor razonable para  $\sigma_y$  es  $\sigma_y = 0.5$ . Este valor asegura una buena capacidad de representación en el área unitaria de la característica visual y un solape despreciable con otras características visuales situadas en otras posiciones [27].

### 3.4.2. Representación de imágenes utilizando el modelo

A la hora de hallar la wavelet mediante la ecuación 3.1.4 usando el modelo analítico tenemos la ventaja de que no tenemos que inferir ninguna información. Dado que estamos usando una función analítica, definida en todo punto, a la hora de evaluar el segundo y controvertido sumando de la ecuación 3.1.4 -la que necesita información fuera del dominio de definición de los datos- no tenemos más que calcular cuánto vale

la función sigmoide en esos puntos fuera del dominio.

Para ilustrar de forma sencilla qué representación de imágenes conseguimos con el modelo analítico hagamos un test similar al que se hizo usando las características visuales empíricas. En el anterior test sólo teníamos 5 filtros diferentes, ahora, sin embargo, al contar con un modelo analítico podemos tener, en principio, tantos como queramos.

Hemos limitado los grados de libertad que definen el banco de características visuales como sigue: primero, hemos fijado algunos parámetros:  $\beta_\infty = 4$  para los tipos  $f_0$ ,  $f_2$ ,  $x_\infty = 0.4$  en el tipo  $f_3$ ,  $x_0 = 0.1$  para el tipo  $f_0$  y  $x_0 = 0.2$  para el tipo  $f_2$ . De tal forma, sólo nos quedan dos parámetros libres:  $\beta$  y el ángulo de orientación de la característica visual (esto es, el ángulo definido por la línea longitudinal). Hemos usado sólo cuatro orientaciones posibles (horizontal, vertical y dos líneas diagonales a  $45^\circ$ ). El parámetro  $\beta$  lo hemos muestreado uniformemente en 10 valores, desde  $\beta = 3$  hasta  $\beta = 4$ . Incluso con estas elecciones modestas en grados de libertad, tenemos un banco con 120 filtros. No pretendemos usar todas las wavelets al mismo tiempo, sólo aquella que está mejor adaptada en cada escala y posición, como hicimos en el caso de las wavelets empíricas.

En las figuras 3.22 y 3.23 presentamos dos ejemplos de imágenes de aproximación  $R_{opt}$  a diferentes resoluciones (esto es, la suma  $\sum_{j=0}^{\infty}$  en la ecuación (3.3.2) es truncada a un máximo  $j$  dado) y las calidades de reconstrucción asociadas (medidas con el PSNR).

Si bien nuestro propósito era formular un modelo que consiguiera describir de manera cualitativa las distintas características visuales agrupándolas en sencillas funciones analíticas, los resultados muestran -con todas sus limitaciones y salvedades- que las imágenes naturales son susceptibles de ser representadas mediante una combinación lineal de distintos filtros obtenidos a partir de un banco analítico de características visuales simples. Los filtros derivados de las características visuales están mejor descritos y tienen un comportamiento más homogéneo al ser calculados mediante una formulación analítica. Eliminar el ruido estadístico presente en el paradigma empírico redundará -como hemos visto- en una mejor representación.

Otra notable diferencia entre el método empírico y el analítico es que en el primero para reconstruir se utilizaba filtros derivados de la estadística de la imagen que se quería

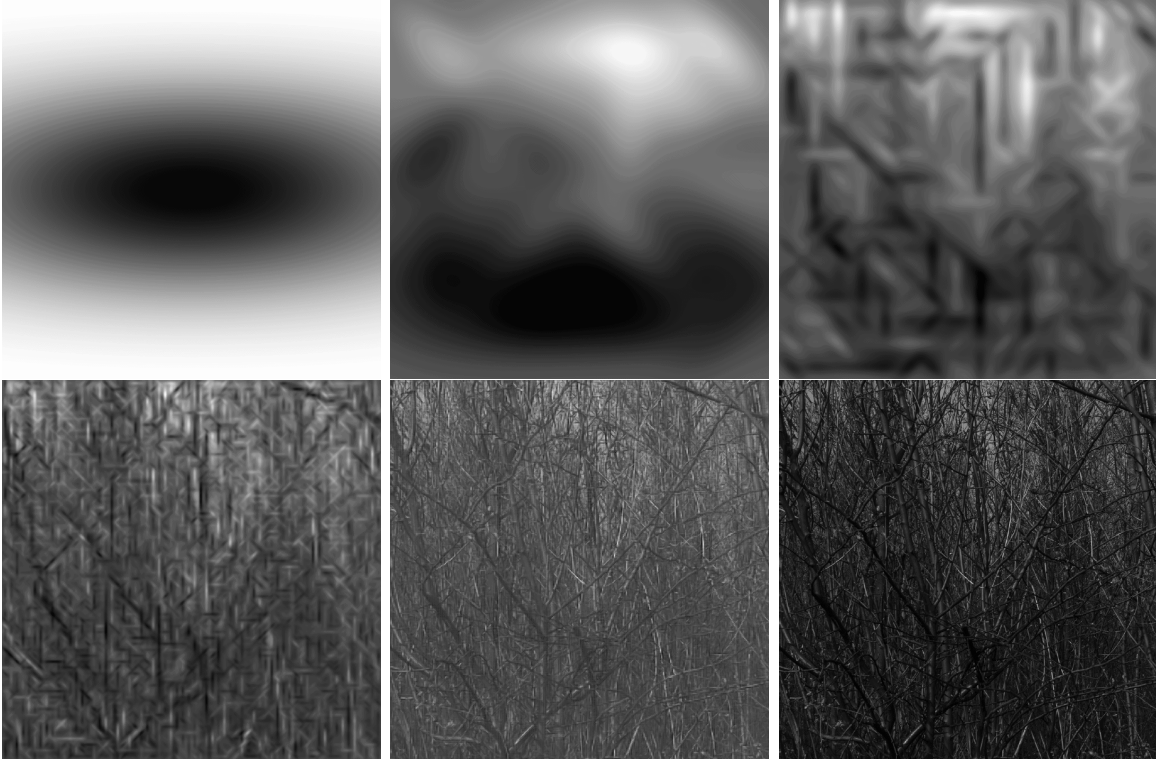


Figura 3.22: **Arriba:**  $R_{opt}$  para varios niveles de resolución. De izquierda a derecha:  $j = 0$  ( $PSNR = 19.70$  dB),  $j \leq 2$  ( $PSNR = 20.50$  dB) y  $j \leq 4$  ( $PSNR = 20.88$  dB). **Abajo:**  $R_{opt}$  para los niveles  $j \leq 6$  (izquierda,  $PSNR = 21.75$  dB),  $j \leq 8$  (centro,  $PSNR = 24.63$  dB) e imagen original *imk00802.imc*.

reconstruir. Sin embargo, en el modelo analítico los filtros son derivados a partir de propiedades estadísticas de -en principio- todo el colectivo. Uno espera que los filtros que se usen ora para descomponer las imágenes naturales, ora como detectores de características visuales, sean lo más universales posibles.

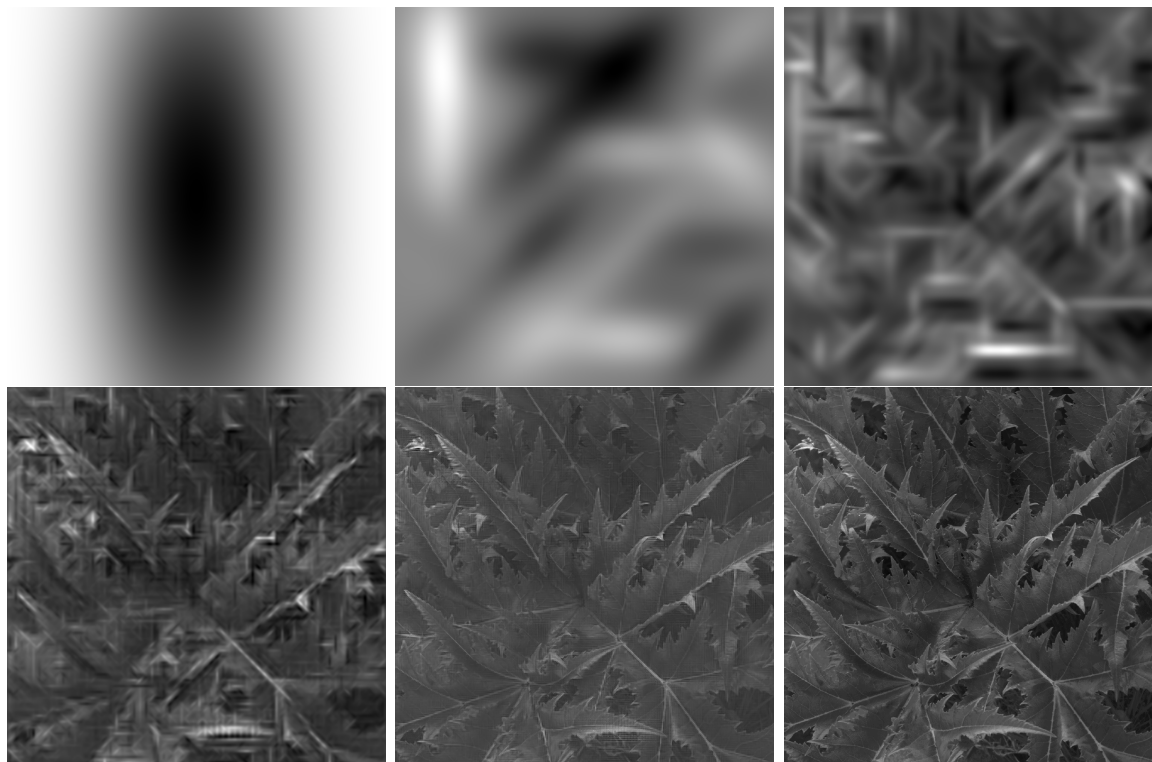


Figura 3.23: **Arriba:**  $R_{opt}$  para varios niveles de resolución. De izquierda a derecha:  $j = 0$  ( $PSNR = 23.02$  dB),  $j \leq 2$  ( $PSNR = 23.74$  dB) y  $j \leq 4$  ( $PSNR = 25.03$  dB). **Abajo:**  $R_{opt}$  para los niveles  $j \leq 6$  (izquierda,  $PSNR = 27.61$  dB),  $j \leq 8$  (centro,  $PSNR = 31.51$  dB) e imagen original *imk01409.imc*.

### 3.5. Características de características visuales: Supercaracterísticas

Motivados por los buenos resultados obtenidos: wavelets extraídas a partir de características visuales con capacidad de reconstrucción, se propuso hacer una clasificación para todo el colectivo de imágenes. En principio, habría que tomar al conjunto de las 4167 imágenes como si de una imagen sola se tratase y clasificar las distintas partes de ésta -por similitud con lo relizado anteriormente: trozos de  $16 \times 16$ -. Por lo tanto, en

este caso, nuestro colectivo a clasificar serían las cerca de 25 millones de subimágenes. Sin embargo, por simplicidad y para rebajar sensiblemente los tiempos de computación se optó en una primer etapa por un análisis taxonómico algo más sencillo.

Del catálogo compuesto por las 4167 imágenes hemos seleccionado de forma aleatoria 500 imágenes a las cuales se les aplicó nuestro algoritmo de clasificación antes descrito. Las 2500 características visuales resultantes de esa taxonomía son mostradas en la figura 3.24. Partamos de estas características visuales que ya hemos podido ver como pueden resumir de manera eficiente al conjunto del colectivo global. Si tomamos esta *imagen* resultante como el conjunto a clasificar, lo que estaríamos haciendo es reducir el colectivo de las 4167 imágenes a una imagen de  $800 \times 800$  píxeles. Como hacíamos con las imágenes individuales, dividiremos ésta imagen *resumen* en trozos de  $16 \times 16$ : las características visuales obtenidas. Aunque hayamos reducido notablemente la estadística, hemos pasado de un escenario donde tenemos que clasificar 25 millones de elementos a otro en donde el número es *sólo* 2500, hay que tener en cuenta el tamaño de las imágenes elegido para su clasificación en comparación al número de características visuales encontradas: 2500 imágenes *promedio* de  $16 \times 16$  píxeles es un conjunto suficientemente representativo.

El solape  $Q$  que se utilizó para la clasificación de las 2500 características visuales fue de 0.7. En este caso se justifica el aumento del parámetro de solape puesto que ahora ya no estamos clasificando imágenes individuales sino características, las cuales definen un espacio de dimensionalidad sensiblemente inferior a 256.

Para ver el decaimiento de las supercaracterísticas se extrajeron las características visuales de cada imagen a la escala doble y para obtener éstas se promedió -con su correspondiente signo- los elementos de cada clase a un escala de  $32 \times 32$  píxeles. Así, las partes centrales de estas características visuales de escala mayor serán las mismas características visuales que ya obtuvimos. No estamos, por tanto, ante una nueva clasificación de subimágenes de tamaño  $32 \times 32$  píxeles pertenecientes a una imagen, pues se ha promediado los elementos que fueron clasificados a la escala de detección:  $16 \times 16$ .

En la tabla 3.1 presentamos las supercaracterísticas extraídas a partir de las 6

primeras clases a la escala de detección y a la escala doble.

### 3.6. Conclusiones

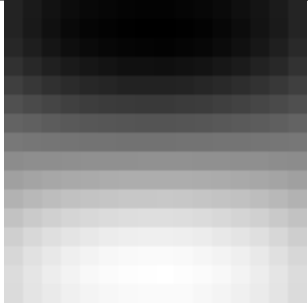
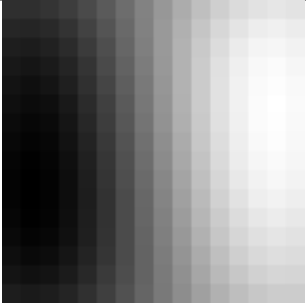
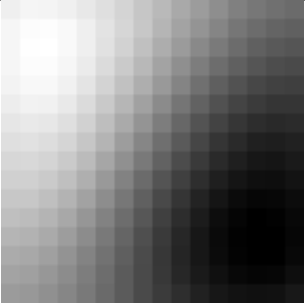
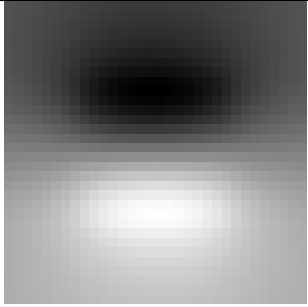
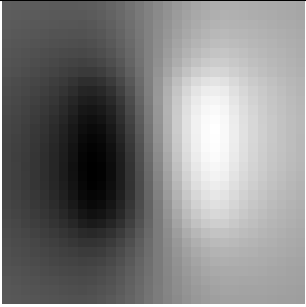
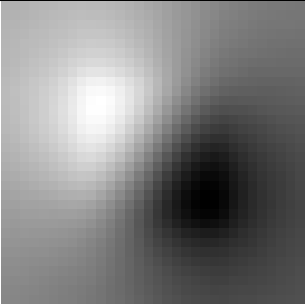
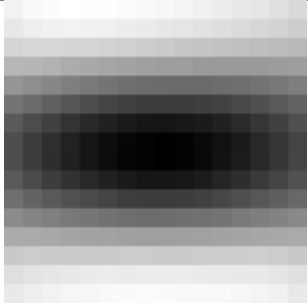
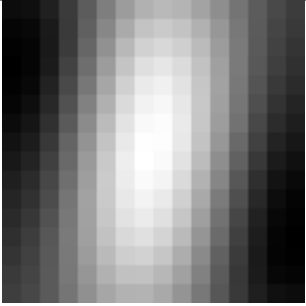
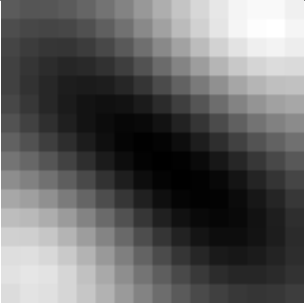
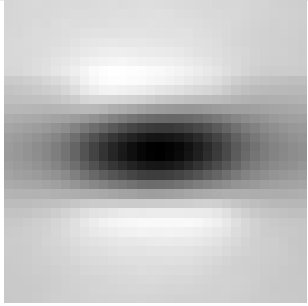
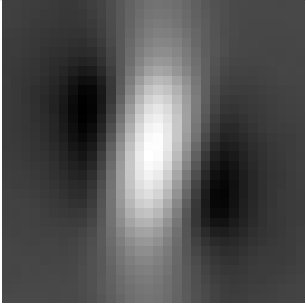
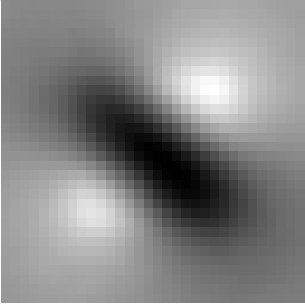
Motivados por la búsqueda de las distintas componentes de la base de wavelet óptima hemos introducido el concepto de característica visual asociado a una determinada voz. Las distintas características visuales han sido extraídas mediante un algoritmo de clasificación de los colectivos de imágenes.

El modelo analítico es una elegante formulación que permite condensar la información del catálogo. Con este modelo conseguimos mejorar sensiblemente la reconstrucción debido en parte a que los filtros derivados son universales.

Si bien el objetivo era construir explícitamente bases de filtros capaces de describir las imágenes naturales, también lo era que éstos condujeran a representaciones de estas imágenes definidas a partir de niveles independientes de resolución (por medio de las variables  $\eta_{rj\vec{k}}$ ).

Como hemos visto, el primer paso en este programa (construcción de bases de representación) es bastante directo. Las dos representaciones propuestas -la basada en los filtros óptimos derivados de las características visuales empíricas y la que parametrizamos con el modelo analítico- parten de un banco de filtros cuyo número es mayor que tres (base sobrecompleta). Es decir, estas bases contienen un número de componentes superior al requerido: en principio, tres. Por tanto, en sentido estricto no son bases. Pero, aun y todo, las imágenes naturales pueden ser reconstruídas por una combinación (una primer etapa lineal seguida de una segunda no-lineal) de ciertos patrones simples presentes en dichas imágenes: las características visuales.

Mientras que la segunda parte (comprobar la condición de independencia) requiere de un estudio matemático específico que excede los límites de esta tesis, y que se propone como trabajo futuro.

29.9 %	21.4 %	4.1 %
		
		
3.0 %	2.8 %	1.4 %
		
		

Cuadro 3.1: Tabla doble: densidad (arriba), supercaracterística visual a la escala de detección (centro) y a escala doble (abajo), ordenadas por sus abundancias relativas dentro del colectivo.

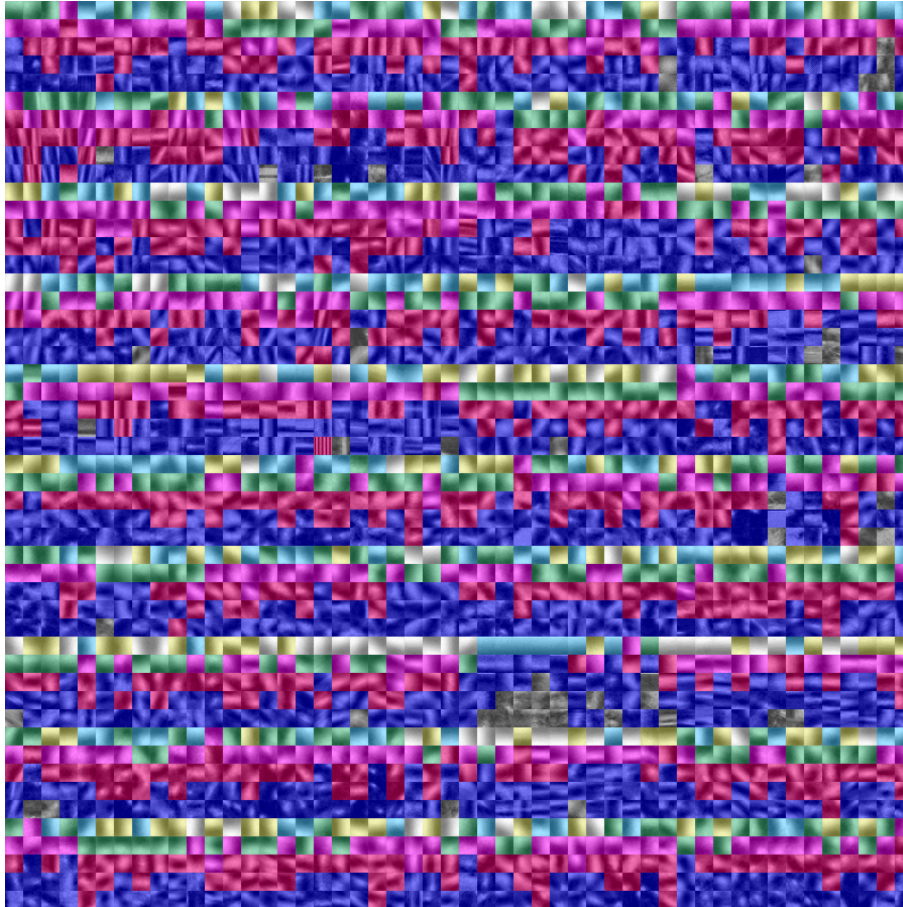


Figura 3.24: Todas las características visuales de las 500 imágenes seleccionadas: esta imagen contiene la información acerca de 2500 características visuales. Cada imagen da lugar a 5 características visuales, que se colocan formando una columna vertical. El nivel de intensidad del brillo de cada pixel es proporcional a la respuesta de la característica visual asociada, brillante para valores positivos y oscuro para negativos. El color sobrepuesto es una indicación del número de subimágenes en la clase, relativo al total de 6144 subimágenes por colectivo (imagen). Se ha usado el siguiente código de colores: 0 a 1 %: Ocre; 1 a 5 %: azul; 5 a 10 %: rojo; 10 a 20 %: magenta; 20 a 30 %: verde; 30 a 40 %: cyan; 40 a 50 %: amarillo; más del 50 %: blanco.



## Capítulo 4

# Revisión de las Hipótesis para Deducir el Filtro Óptimo

Acabábamos el capítulo anterior aplicando el algoritmo de clasificación de características visuales a un colectivo de *imágenes* que eran, a su vez, características visuales, y como resultado de esta clasificación obtuvimos las supercaracterísticas visuales. Al observar las distintas clases uno observa lo regulares que son éstas y el enorme parecido que tienen a un tipo de wavelet que se usa como analizador: la función de Gabor, función que se suele usar para parametrizar los campos receptivos, [36]. Esto nos hace plantearnos si en este caso la fórmula de obtención del filtro óptimo no adopta una forma diferente a la planteada por la ecuación 3.1.4 y lo que nosotros llamamos supercaracterísticas visuales no son directamente las distintas componentes de una posible base de wavelets óptimas.

La motivación para hacer una clasificación de imágenes fue la búsqueda de una solución al problema de las voces: con una única wavelet no eramos capaces de reconstruir de manera eficiente las imágenes del colectivo. La wavelet obtenida a partir del colectivo de las 4167 imágenes no era suficiente para poder representar las diferentes imágenes, por ser ésta -en principio- una mezcla de las diferentes voces que componen una base de wavelet. Se buscó entonces algún método que pudiera descomponer las distintas voces o componentes que componían esa wavelet mezcla.

A la hora de aplicar la fórmula de obtención de la wavelet óptima, ecuación 3.1.4, una vez extraídas las diferentes clases, no usamos, como era preceptivo, la media de las imágenes del colectivo de la clase. Argumentando propiedades de simetría utilizamos el concepto de media signada para así potenciar la característica visual del colectivo, y esa media signada es la que usamos para calcular posteriormente el filtro. De la misma manera, asociamos el filtro obtenido a partir de una clase concreta de dicho colectivo como una de las voces o componentes de una de las posibles bases de wavelets óptimas.

En la deducción original de la fórmula de la wavelet no se exploró de forma rigurosa el impacto que podía tener los signos de las variables aleatorias  $\alpha$  y  $\eta$ . Veremos como un estudio detallado podrá dar una explicación al rango de aplicación de la fórmula utilizando medias signadas y nos introducirá un nuevo grado de libertad. También, nos permitirá afirmar que el método de clasificación de imágenes dentro de un colectivo es una forma de extracción de las componentes de una base de wavelets.

Si bien parte de estas cuestiones ya las hemos abordado, los argumentos que se han dado han sido meramente intuitivos. Por ello es necesario hacer una revisión en profundidad de la derivación de la fórmula de obtención del filtro óptimo para poder, por un lado, dar solidez matemática a los argumentos intuitivos y, por otro, dar respuesta a las cuestiones anteriormente planteadas.

## 4.1. Estudio de los signos

### 4.1.1. Preliminares

Hemos visto que la wavelet óptima (o la mezcla, si se trata de varias voces) asociada a un determinado colectivo de imágenes no es más que la diferencia del promedio del contraste en dos escalas consecutivas. Sin embargo, siendo rigurosos, a esa diferencia había que haberla multiplicado por un factor:  $1/\langle\alpha_{0\vec{0}}\rangle$ , si bien, como ya explicamos, este factor no era en principio importante pues existía una libertad: si  $\Psi$  es una wavelet óptima también lo era  $a\Psi$  (con  $a$  real distinto de cero). Con la elección de la normalización restringíamos tal libertad. Pero, ¿qué ocurre si  $\langle\alpha_{0\vec{0}}\rangle = 0$ ? Es decir, que

el promedio del contraste dentro del colectivo es una fluctuación entorno a cero. No podríamos despejar la wavelet como lo hemos hecho. En los casos anteriores (4167 imágenes y características visuales) en los que la hemos utilizado, ¿se ha verificado que es distinto de cero?

Para intentar resolver estas cuestiones hemos realizado una serie de cálculos experimentales utilizando como banco de datos un subcolectivo de 1000 imágenes ( $1024 \times 1024$ ) de van Hateren.

Para realizar este estudio hemos utilizado seis bases de wavelet diádicas. Las seis bases consideradas son las bases separables 2D de Haar, Daubechies  $p = 2$ , Daubechies  $p = 4$  y Battle-Lemarié (splines) de orden 1, de orden 3 y de orden 6. No consideramos en este análisis la wavelet óptima obtenida experimentalmente a partir de las 4167 imágenes porque es justo lo que queremos revisar: la fórmula de obtención de ésta a partir de un colectivo de imágenes naturales, además hay que tener en cuenta que -como ya vimos en el capítulo 2- dicha wavelet es una mezcla de las diferentes voces -con la voz horizontal como un peso mucho mayor- que no nos permite discriminar en orientación -ni tampoco sabemos cómo-. Para cada una de las bases mencionadas se calculó la estadística de los signos de los  $\alpha_{00}$  para las distintas voces Horizontal, Vertical y Diagonal. Las wavelets utilizadas no son óptimas (existencia de unas variables  $\eta$ 's independientes) pero es versosímil que el estudio pueda ser extrapolado a la base óptima.

Es evidente que el estudio que hacemos para la escala global se puede extrapolar, en principio, para cualquier otra escala y posición debido a la simetría de invariancia de escala que presentan las imágenes naturales.

#### 4.1.2. Signo de $\alpha$ : $\sigma$

En el cuadro 4.1 mostramos los resultados obtenidos: son bastante elocuentes. El comportamiento de los  $\alpha$ 's es muy diferente según qué dirección analicemos. Para los casos Vertical y Diagonal los signos aparecen esencialmente equidistribuidos, sin embargo, la dirección Horizontal presenta un desequilibrio muy bien definido en los signos a favor de los positivos (o de los negativos en el caso de Daubechies  $p = 4$  y Battle-

	$\%(\alpha_H > 0)$	$\%(\alpha_V > 0)$	$\%(\alpha_D > 0)$
<b>Haar</b>	84.30	51.40	47.40
<b>Dau p=2</b>	64.80	53.00	46.50
<b>Dau p =4</b>	18.60	49.70	51.50
<b>B-M 1</b>	75.00	50.40	52.00
<b>B-M 3</b>	74.40	50.10	52.50
<b>B-M 6</b>	17.00	48.30	48.30

Cuadro 4.1: Distribución de los signos de los  $\alpha_{0\bar{0}}$  (expresados en porcentaje) de las proyecciones en seis bases de wavelets y para un subcolectivo de 1000 imágenes de van Hateren. Las etiquetas  $H$ ,  $V$  y  $D$  corresponden, respectivamente, a las wavelets Horizontal, Vertical y Diagonal, respectivamente

Lemarié de orden 6) que puede ser cuantificado alrededor de una proporción 80 %-20 %.

A la vista de estos resultados nos es fácil comprender por qué en el colectivo de las 4167 imágenes se obtuvo un filtro con una orientación muy definida. Como punto de partida ya observamos que su promedio nos da un resultado distinto de cero. Pero además, al promediar las imágenes de dicho colectivo nos encontramos que para cada característica visual (cuyos filtros asociados les hacemos corresponder con las diferentes voces) las contribuciones en los dos *sentidos* son similares, sin embargo hay una determinada donde existen más contribuciones en uno de los sentidos: el borde horizontal. Este sesgo en el signo de la característica horizontal pone en evidencia la existencia de un gradiente homogéneo en el contraste a la escala global. Como ya avanzamos, el que en media nos encontremos con que haya más luz arriba que abajo en las imágenes proviene del hecho de que exista un foco de luz principal que es común en todas éstas y está situado en la parte superior: el sol. Así pues, debido a la existencia de esta asimetría en la fuente de luz en el colectivo estudiado pudimos obtener una media (distinta de cero) que en modo efectivo corresponde con una sola característica visual. A partir de ésta se pudo calcular su filtro asociado.

Hemos de entender, por tanto, que el conjunto de las imágenes de van Hateren

tienen una elección de orientación -ruptura de la isotropía- por el modo en que las tomó: la cámara con la que tomó los datos siempre tuvo una orientación horizontal. Esta anisotropía que se ve reflejada sin ningún genero de dudas a la escala global; a otras escalas más finas, aunque presente, se irá difuminando.

Podríamos preguntarnos si el filtro obtenido a partir de ese colectivo es un filtro *puro* o, por el contrario, se trata de una combinación lineal de varios, asociados cada uno a su característica visual. Veremos que en realidad es las dos cosas, dependiendo qué base utilicemos.

Hasta ahora hemos descrito el escenario que se tiene cuando se considera un colectivo con sesgo. Veamos ahora de manera rigurosa la asociación que hacemos entre característica y elemento de una base de wavelet, así como el método de extracción del filtro óptimo para colectivos con media nula: qué significa signar y por qué eso nos sirve para clasificar.

## 4.2. Medias signadas

Partamos de que el contraste de una imagen lo podemos descomponer utilizando una determinada base de wavelets para así llegar a la expansión ya conocida:

$$c(\vec{x}) = \sum_{r=0}^{n-1} \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathbb{Z}_{2^j})^2} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x}) . \quad (4.2.1)$$

Multipliquemos ambos lados de la expresión por el signo del coeficiente a la escala más grande y de componente  $r = s$ , esto es, el signo de  $\alpha_{s0\vec{0}}$  que denotaremos por  $\sigma_{s0\vec{0}}$ . Tomemos el valor esperado,

$$\langle \sigma_{s0\vec{0}} c(\vec{x}) \rangle = \sum_{r=0}^{n-1} \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathbb{Z}_{2^j})^2} \langle \sigma_{s0\vec{0}} \alpha_{rj\vec{k}} \rangle \Psi_{rj\vec{k}}(\vec{x}) . \quad (4.2.2)$$

Si hacemos uso, una vez más, del proceso multiplicativo y de la independencia de las variables  $\eta$

$$\langle \sigma_{s0\vec{0}} c(\vec{x}) \rangle = \sum_{r=0}^{n-1} \langle \sigma_{s0\vec{0}} \alpha_{r0\vec{0}} \rangle \sum_{j=0}^{\infty} \sum_{\vec{k} \in (Z_{2j})^2} \langle \eta_r \rangle^j \Psi_{rj\vec{k}}(\vec{x}) . \quad (4.2.3)$$

Analicemos ahora con detalle el término

$$\sum_{r=0}^{n-1} \langle \sigma_{s0\vec{0}} \alpha_{r0\vec{0}} \rangle = \langle |\alpha_{s0\vec{0}}| \rangle + \sum_{r \neq s} \langle \sigma_{s0\vec{0}} \rangle \langle \alpha_{r0\vec{0}} \rangle . \quad (4.2.4)$$

Suponiendo que existe independencia entre  $\sigma_{s0\vec{0}}$  y  $\alpha_{r0\vec{0}}$  (es decir, haciendo la hipótesis de que, a una escala y posición dadas, los signos de los  $\alpha$  son independientes si las componentes son distintas) nos podremos quedar sólo con el primer término  $\langle \alpha_{s0\vec{0}} \rangle$  si el valor esperado de  $\sigma_{s0\vec{0}}$  o el de  $\alpha_{r0\vec{0}}$  es cero.

Como hemos podido ver anteriormente, dependiendo del colectivo elegido -o clasificado- el comportamiento de los  $\alpha$ 's será diferente según qué voz analicemos. Si la voz elegida para signar,  $r = s$ , no tiene un sesgo en los signos -a esa escala-, entonces podremos quedarnos con el primer término. Puede pasar, sin embargo, que exista un sesgo en  $\sigma_{s0\vec{0}}$  y su valor esperado no sea cero -o, si tratamos con promedios, no tienda a cero a medida que aumentamos la estadística-. Entonces deberemos concentrarnos en analizar el término  $\langle |\alpha_{r0\vec{0}}| \rangle$  para todos los  $r \neq s$ .

Tanto las imágenes tomadas por van Hateren como los colectivos de imágenes a los que se les aplicó el algoritmo de clasificación, tienen un sesgo inducido en una determinada voz a una escala. Esto es resultado de la manera en que se fueron tomando las imágenes: la cámara digital sostenida en la cabeza estuvo siempre en la misma orientación horizontal, provocó que a la escala global hubiera una dirección privilegiada y, dentro de ésta, también un sentido. A la hora de clasificar las subimágenes estamos seleccionando individuos con una relación de equivalencia muy determinada; además, el hecho de signar dichos individuos provoca una sobrerrepresentación en uno de los sentidos.

Pero para colectivos de imágenes sin un sesgo inducido -ya sea en la forma de medir o por una preselección de los individuos- es lógico pensar que el valor esperado de los  $\alpha$  a una escala y posición determinada sea típicamente cero para todas las voces por

la invariancia de escala y de traslación. Lo que nos conduciría a que en estos colectivos el sumatorio de la ecuación 4.2.4 es cero y sólo sobreviviría el primer término.

Podríamos, por último, encontrarnos con un último caso hipotético y algo rebuscado: un colectivo en el que hubiera un sesgo en el signo -y a la escala global- en una voz distinta -o, incluso, en más de una- de la que hemos *escogido* para signar y que esta componente *también* presenta un sesgo en el signo a esa misma escala global. En este caso podríamos redefinir una nueva base de wavelets que tuviera una componente que fuera una combinación lineal adecuada de todas esas voces con sesgo de la antigua base. De esta forma, estaríamos ante el mismo escenario del párrafo anterior.

En resumen, la expansión del promedio de las imágenes signadas -lo que, por definición, venimos llamando característica visual- se simplifica una vez que incluimos las consideraciones anteriores

$$\langle \sigma_{s0\vec{0}} c(\vec{x}) \rangle = \langle |\alpha_{s0\vec{0}}| \rangle \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathbb{Z}_{2^j})^2} \langle \eta_s \rangle^j \Psi_{sj\vec{k}}(\vec{x}) . \quad (4.2.5)$$

Por el hecho de haber signado resolvemos el problema del signo de  $\alpha$  pero, debido a que cada característica visual involucra sólo una única wavelet -ya no tenemos una mezcla, como pasaba anteriormente-, además conseguimos separar las diferentes voces de la base de wavelet. Ya no necesitamos imponer, como decíamos en el anterior capítulo, la existencia de una voz que sea la dominante para toda escala y posición dentro del colectivo escogido: estábamos eligiendo un colectivo de imágenes muy particular donde existía una voz que persistía a las diferentes escalas y posiciones. Los promedios signados se convierten, entonces, en selectivos a cada voz. Por ello, podemos hacer corresponder cada característica visual con una componente de una base de wavelets.

En todo el desarrollo anterior se ha supuesto el conocimiento *a priori* del signo de los  $\alpha$  o, si se quiere, de al menos el del signo del escogido para signar. Sin embargo, para conocer el signo de los  $\alpha$  necesitamos forzosamente saber antes cuál es la base de wavelets óptima. Pero eso es lo que estamos intentando hacer, encontrar la fórmula de obtención de las componentes de la base óptima de wavelets. Necesitamos, en consecuencia, un estimador de esos signos,  $\sigma_{s0\vec{0}}$ . En el capítulo anterior, motivados por la necesidad de aumentar la estadística de las características visuales, a cada imagen de

una determinada clase se le asignó un signo que venía definido por el producto escalar de dicha imagen con la imagen central de la clase. De tal forma que un cambio el signo no era más que una inversión en el gradiente de luminosidad. ¿Podría este signo así definido ser un buen estimador? En verdad, no necesitamos tanto, de hecho necesitamos *algo más que suerte*. Lo que se le tiene que imponer es que la probabilidad de acierto -o desacierto- sea mayor que 1/2:

$$P(\epsilon_{s0\vec{0}}\alpha_{s0\vec{0}}) \equiv P_{\epsilon_{s0\vec{0}}} \neq \frac{1}{2}, \quad (4.2.6)$$

donde  $\epsilon_{s0\vec{0}}$  es el estimador de  $\sigma_{s0\vec{0}}$ . Es decir, necesitamos que el estimador elegido esté correlado con el el signo: que no sea *simple azar*.

El uso de estimadores introduce una modificación en la ecuación 4.2.5

$$\langle \epsilon_{s0\vec{0}} c(\vec{x}) \rangle = (2P_{\epsilon_{s0\vec{0}}} - 1) \langle |\alpha_{s0\vec{0}}| \rangle \sum_{j=0}^{\infty} \sum_{\vec{k} \in (\mathcal{Z}_{2^j})^2} \langle \eta_s \rangle^j \Psi_{sj\vec{k}}(\vec{x}). \quad (4.2.7)$$

## 4.3. Persistencia de los signos

### 4.3.1. Preliminares

En la deducción de la fórmula de la wavelet óptima se necesitó hacer dos hipótesis acerca de la variable independiente  $\eta$ . Primero, independencia estadística entre el valor absoluto de  $\eta$  y su signo (que llamaremos  $\beta$ ). Segundo, debido a que la invariancia de traslación imponía que el valor esperado de  $\beta$  fuese cero, se supuso que el promedio de los signos venía gobernado por las fluctuaciones. La independencia entre el módulo y el signo está verificada experimentalmente, [29]. Sin embargo, aunque la segunda de las hipótesis veremos que en verdad no es necesaria (la invariancia de traslación no se rompe con un valor esperado de  $\beta$  distinto de cero), el estudio de la distribución de los signos de  $\beta$  nos muestra que su promedio es mayor que cero y que, por tanto, existe *también* persistencia en el signo de los  $\alpha$  en dos escalas consecutivas.



### 4.3.2. Signo de $\eta$ : $\beta$

A imagen del estudio de los signos de los  $\alpha$ , hemos analizado el comportamiento de los signos de los  $\eta$  a diferentes escalas (promediando los distintos  $\vec{k}$ ). Empero, en esta ocasión, sólo se ha podido utilizar una sólo wavelet: la óptima, dado que las variables  $\eta$ , definidas como el cociente entre los  $\alpha$  a dos escalas consecutivas, son independientes si estamos trabajando con la wavelet óptima. Otra diferencia es que el colectivo usado en esta ocasión es todo el conjunto de las 4167 imágenes.

En el cuadro 4.2 se resume el análisis efectuado para varias escalas. La escala  $j = 0$  de la tabla es especial, ya que en realidad se corresponde con los valores obtenidos a partir de  $\alpha_{0\vec{0}}$ , no de  $\eta_{rj\vec{k}}$ , que no está definido a esa escala. Esa línea de la tabla nos muestra un resultado parecido a lo que se obtuvo con las otras wavelets: aparece un desequilibrio en los signos (en este caso, también a favor de los negativos) entorno a 40 % – 60 %.

Se ha introducido una nueva variable  $\lambda \equiv \bar{\beta}|\bar{\eta}|$  que mide el grado de independencia entre el signo y el módulo de las variables estudiadas. Para poder confirmar la independencia entre el signo de  $\eta$  y su módulo se ha calculado también éste último para poder evaluar  $\lambda$ , así como la variable  $\eta$ . El grado de proximidad entre los dos valores,  $\eta$  y  $\lambda$ , en todas las escalas mostradas (obviamente, no debe mirarse la escala  $j = 0$ ) es una confirmación también experimental de la independencia entre el signo y módulo de  $\eta$  verificada en [29]. Por último, se observa que la independencia entre estas dos variables es casi perfecta para la escala  $j = 6$ . Esto puede ser debido a que en esta escala tenemos suficiente estadística, el número de valores de  $\eta$  que promediamos a una escala es cuatro veces mayor que a la escala anterior debido al vector de posición,  $\vec{k}$ ; sin llegar al límite de resolución de la imagen ( $j = 9$ ) que introduciría ruido por *el efecto de frontera*.

Centrándonos en el promedio de los signos de  $\eta$  observamos como para todas las escalas examinadas su valor es claramente mayor que cero. Hay una persistencia estadística en el signo de  $\alpha$  durante las escalas. Esto implica que si uno coge una imagen del colectivo examinado y a una escala cualquiera evalúa el signo de la proyección en wavelet,  $\sigma$ , el valor de éste (tanto da si es positivo como si es negativo) será el que nos

encontremos con mayor probabilidad a la siguiente escala.

	$\bar{\eta}_j$	$\bar{\beta}_j$	$ \bar{\eta}_j $	$\lambda_j$
$j = 0$	-35.80	-0.38	52.76	-20.05
$j = 2$	0.07	0.16	0.71	0.11
$j = 4$	0.05	0.11	0.71	0.08
$j = 6$	0.21	0.39	0.51	0.20
$j = 8$	0.32	0.57	0.43	0.25

Cuadro 4.2: Promedios de los  $\eta$ , de sus signos ( $\beta$ ), de sus módulos ( $|\eta_j|$ ) y de los  $\lambda_j$  para distintas escalas para el colectivo de las 4167 imágenes.

Sin embargo, podemos encontrar otros colectivos de imágenes donde hubieramos obtenido un valor de  $\bar{\beta}$  compatible con cero, que implicaría que no hay persistencia: no podemos predecir cómo va a ser el signo de  $\alpha$  de una imagen a una escala dada sabiendo el signo a una escala más grosera. Seguramente esto es lo que esté pasando para el colectivo de las características visuales cuya clasificación nos dio como resultado las supercaracterísticas. En cualquier caso, es un estudio que hemos de verificar.

Finalmente, para un colectivo de imágenes con  $\bar{\beta} < 0$  tendríamos un fenómeno de antipersistencia: los signos de los  $\alpha$  de una imagen dada (de dicho colectivo) a dos escalas consecutivas serían con mayor probabilidad opuestos.

Al igual que pasaba con el signo de los  $\alpha$ , para conocer el promedio de los signos de los  $\eta$ ,  $\bar{\beta}$ , fundamental para derivar la fórmula de obtención del filtro óptimo, necesitamos saber antes cuál es esta base de wavelets, que dependerá del colectivo de imágenes. La ignorancia primigenia del valor de este promedio que cuantifica la persistencia en el signo de  $\alpha$  habremos de parametrizarla con una variable que estará comprendida entre los puntos extremos que este promedio puede tomar:  $-1$  y  $1$ . Sólo *a posteriori* podremos verificar si el valor escogido para  $\bar{\beta}$  es coherente con el colectivo de imágenes que hemos utilizado para derivar el filtro óptimo.

### 4.3.3. Análisis de la persistencia

Para introducir la dependencia que tiene el signo de los  $\eta$  en la fórmula de obtención de la wavelet óptima, así como desarrollar de forma explícita el fenómeno de la persistencia analicemos con más rigor el valor medio de  $\eta_{rj\vec{k}}$ . Esta variable aleatoria toma valores positivos y negativos por lo que podemos factorizarla de la siguiente forma,

$$\eta_{rj\vec{k}} = |\eta_{rj\vec{k}}| \beta_{rj\vec{k}}, \quad (4.3.1)$$

siendo el valor absoluto y el signo dos variables aleatorias, la primera continua y la segunda de ellas discreta y que sólo toma dos valores: -1,+1. Si ahora tomamos el valor medio obtenemos,

$$\langle \eta_{rj\vec{k}} \rangle = \langle |\eta_{rj\vec{k}}| \rangle \langle \beta_{rj\vec{k}} \rangle + \mu(|\eta_{rj\vec{k}}| \beta_{rj\vec{k}}), \quad (4.3.2)$$

donde  $\mu(|\eta_{rj\vec{k}}| \beta_{rj\vec{k}})$  denota la covariancia entre las dos variables. Dado que todos los  $\eta_{rj\vec{k}}$  poseen la misma distribución su valor medio no puede depender de  $j$  ni  $\vec{k}$ , por otra parte también es sabido que el promedio del valor absoluto tampoco depende de la escala y posición. Si hacemos la hipótesis ya discutida con anterioridad de que el signo de  $\eta_{rj\vec{k}}$  es independiente del valor absoluto, se llega a la conclusión de que el valor medio del signo no puede depender de los dos índices,  $j, \vec{k}$ , por lo tanto, en particular, el promedio de los signos respeta la invariancia de traslación. Por tanto la ecuación se simplifica,

$$\langle \eta_r \rangle = \langle |\eta| \rangle \langle \beta_r \rangle. \quad (4.3.3)$$

Esta última relación invalida la segunda de las hipótesis utilizada para la deducción de la wavelet óptima porque cualquier valor que tome el valor esperado de los signos mantiene la invariancia traslacional.

Cuando pasamos de una escala a otra sabemos, dada la relación multiplicativa 2.4.5 y la esparsidad en la distribución de los  $\eta_{rj\vec{k}}$ , que va a existir una persistencia en el código definido por los  $\alpha_{rj\vec{k}}$ . Si nos fijamos sólo en el signo de estos coeficientes tendremos

$$\sigma_{rj\vec{k}} = \beta_{rj\vec{k}} \sigma_{rj-1[\frac{\vec{k}}{2}]} . \quad (4.3.4)$$

Esta ecuación nos dice que para que el signo de  $\alpha_{rj\vec{k}}$  *persista* (sea igual al de  $\alpha_{rj-1[\frac{\vec{k}}{2}]}$ )  $\beta_{rj\vec{k}}$  debe tomar el valor  $+1$ . Aún existiendo esta persistencia, habrá veces que se introduzca alguna novedad en el signo ( $\sigma_{rj\vec{k}}$  tendrá signo opuesto a  $\sigma_{rj-1[\frac{\vec{k}}{2}]}$ ), en este caso, que será menos probable que el anterior,  $\beta_{rj\vec{k}}$  valdrá  $-1$ .

#### 4.4. Impacto de los signos en la fórmula de obtención del filtro óptimo

Una vez analizado el signo de  $\eta$ , la fórmula 2.5.4 queda sutilmente modificada

$$\Psi_r(\vec{x}) = \frac{1}{\langle |\alpha_{r0\vec{0}}| \rangle (2P_{\epsilon_{r0\vec{0}}} - 1)} \left( \langle \epsilon_{r0\vec{0}} c(\vec{x}) \rangle - \lambda_r \sum_{\vec{k} \in (\mathbb{Z}_2)^2} \langle \epsilon_{r0\vec{0}} c(2\vec{x} - \vec{k}) \rangle \right). \quad (4.4.1)$$

Las dos fórmulas son muy parecidas: en vez del promedio del módulo de  $\eta$  aparece simplemente  $\eta$ . Será válida siempre que  $P_{\epsilon_{r0\vec{0}}} \neq 1/2$  e introducimos  $\lambda_r \equiv \langle \beta_r \rangle \langle |\eta| \rangle$ : el nuevo parámetro que tiene en cuenta el signo de  $\eta_r$ , que estará comprendido entre  $-1/2$  y  $1/2$ .

Con la nueva formulación no necesitamos que el promedio de  $\alpha_{r0\vec{0}}$  sea distinto de cero, es válida para todos los colectivos de imágenes, tengan éstos o no algún sesgo. Además, el hecho de signar no introduce ninguna modificación en el colectivo, es decir, que si un colectivo tuviere un sesgo determinado -como el de las 4167 imágenes- el que se signe no debería introducir ningún cambio en  $\lambda$ .

En el caso particular de no haya persistencia,  $\lambda = 0$ , de manera automática las características visuales son las wavelets. Este puede ser el caso de las supercaracterísticas visuales hayadas en el anterior capítulo. Como las supercaracterísticas visuales son obtenidas de clases más globales (se mezclan características visuales semejantes de muchas imágenes) es natural pensar que para éstas no haya persistencia en el signo de  $\alpha$ . Las formas observadas en la escala  $32 \times 32$  indican que las supercaracterísticas visuales podrían razonablemente ser las wavelets. Aunque eso no excluye que para cada clase en cada imagen sea necesario considerar un  $\lambda$  diferente de cero.

## 4.5. Conclusiones

La nueva fórmula no es susceptible al promedio del signo de  $\alpha$ : esto es, siempre es válida. Sin embargo, la formulación antigua no será aplicable en todos los supuestos, y aún cuando se verifiquen las condiciones de aplicabilidad (que tiene en cuenta el valor medio de los  $\alpha$ 's a la escala global) requerirá de unas condiciones mucho más estrictas que el modelo de las medias signadas. En cualquier caso, para hallar la wavelet óptima debemos asumir *a priori* un valor para el promedio del signo de  $\eta$ , y que, en principio, dependerá del colectivo de imágenes. Hemos de introducir un parámetro cuyo valor está comprendido entre  $-1$  y  $1$ .

Al utilizar los promedios signados también conseguimos separar las diferentes voces de la base de wavelet. No es necesario imponer la existencia de una voz que sea la dominante para toda escala y posición dentro del colectivo escogido. Los promedios signados se convierten en selectivos a cada voz.

Un procedimiento para encontrar el valor de  $\beta$  asociado a un colectivo de imágenes dado consistiría en un método iterativo que buscara la convergencia. Partiendo de un valor inicial para el promedio del signo de  $\eta$  calculamos la wavelet a partir de 4.4. Una vez que tenemos la base de wavelets hallamos qué valor de  $\beta$  obtenemos para el colectivo. Este nuevo valor (si no coincide con el inicial) lo usamos para calcular otra vez la wavelet y seguir iterativamente hasta que consigamos converger a un valor concreto.

Aún no teniendo resultados concluyentes para el valor medio de  $\alpha_{rj\vec{k}}$ , en el próximo capítulo veremos como los filtros derivados con  $\lambda$  compatible con cero (que es lo mismo que decir que no hay persistencia en el módulo del signo de  $\eta$ ) son los que mejor se ajustan a funciones de Gabor: funciones prototípicas para parametrizar los campos receptivos.



## Capítulo 5

# Campos Receptivos de Células Simples

Es sabido que la adaptación de los seres vivos a su entorno afecta al código sensorial de dos formas complementarias: minimizando la redundancia en el código que codifica la información contenida en los estímulos, [37], y maximizando la información de dicho código respecto al estímulo, [3], si bien sabemos que ambas estrategias están estrechamente relacionadas, [4]. Los distintos modelos propuestos que se utilizan para predecir el código sensorial tienen en común la necesidad de adoptar un tratamiento estadístico de los estímulos sensoriales. En particular, el sistema visual *tiene* que usar estas estrategias (que codifique y compactifique la información proveniente de los estímulos) para poder *comprender* de manera eficiente el medio.

Hemos estudiado a lo largo de los capítulos precedentes las propiedades estadísticas de las imágenes naturales para poder deducir un código óptimo que posteriormente nos permita *decir algo* sobre el comportamiento del sistema visual: el conocimiento de la estadística de las imágenes naturales permite hacer predicciones sobre propiedades de células sensoriales visuales, [6]. A partir de todas estas premisas el código sensorial debe de haber aprendido, por tanto, después de la exposición continuada a los estímulos visuales -las imágenes naturales- a lo largo de su aprendizaje, que existen ciertas propiedades -estadísticas- en las imágenes naturales.

En este capítulo presentaremos un modelo que permite deducir la estructura de los campos receptivos de las células simples a partir de los dos principios que han emergido del estudio de las imágenes naturales: existencia de un código óptimo y posibilidad de reconstrucción de éstas a partir de patrones simples. Justamente, van a ser los filtros asociados a las distintas características visuales los candidatos a predecir la forma de los campos receptivos.

Para poder asociar el filtro que obtenemos a partir de las características visuales con los campos receptivos debemos hacer una descripción biológica de la fisiología de la neurona y del sistema visual primario. Del potencial de acción a su respuesta lineal para poder llegar a los campos receptivos.

En la literatura existen diversos modelos que *asumiendo a priori* propiedades estadísticas de los estímulos visuales intentan predecir la estructura de los campos receptivos: [37],[6],[8],[10],[38]. Sin embargo, veremos como estas predicciones de los campos receptivos no consiguen explicar de manera satisfactoria los resultados observados, [39].

## 5.1. Del potencial de acción a los campos receptivos

### 5.1.1. Potencial de acción

Una neurona puede emitir una señal eléctrica relevante para el sistema nervioso debido a que existe una diferencia de potencial eléctrico entre el interior de la célula y el exterior. En condiciones de reposo, el voltaje dentro de la neurona es del orden de  $-70mV$  respecto del líquido extra-celular (que, por conveniencia, se fija allí a  $0mV$ ), se dice entonces que la neurona está polarizada. Bombas iónicas localizadas a lo largo de toda la membrana mantienen gradientes en la concentración para compensar esa diferencia de potencial eléctrico. Por ejemplo, los iones de sodio están mucho más concentrados fuera de la neurona que dentro, sin embargo la concentración de los iones de potasio es sensiblemente inferior en el medio extra-celular que en el interior de la neurona. Los iones se mueven hacia dentro y fuera debido al gradiente en la concentración y la diferencia de voltaje. Cuando los iones positivos salen del interior



de la célula hacia el exterior, o bien cuando iones negativos entran en la neurona a través de canales existentes en la membrana, la neurona se hiperpolariza; mientras que si ocurre lo contrario la neurona se depolariza.

Si una neurona se despolariza lo suficiente para que la diferencia de potencial alcance un cierto umbral, entonces se desencadena un proceso de realimentación que conducirá a que la neurona produzca un potencial de acción. Un potencial de acción (también se llamará spike o que la neurona ha emitido un disparo) es una fluctuación de unos  $100mV$  en la diferencia de potencial eléctrico con una duración de alrededor de  $1ms$  (figura 5.1). La generación del potencial de acción depende muy fuertemente de si ha habido un potencial anterior. Unos pocos milisegundos después de producirse un spike es prácticamente imposible que se vuelva a producir otro. Este periodo se le suele conocer como periodo refractario absoluto. Por un periodo que se produce alrededor de  $10ms$  después del disparo, es muy difícil que la neurona dispare. A este otro periodo se le llama periodo refractario relativo.

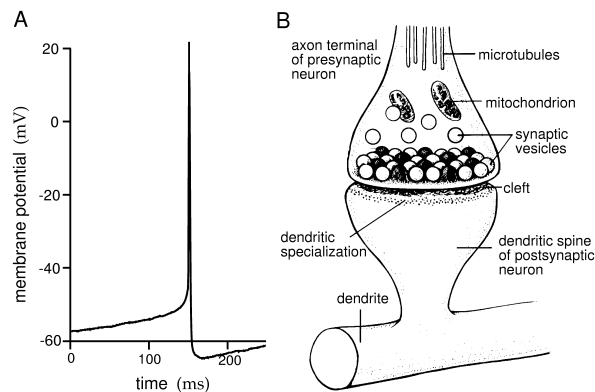


Figura 5.1: A. Registro intracelular de un potencial de acción de una neurona cortical de una rata. B. Diagrama de una sinapsis.

Los potenciales de acción son muy importantes porque son los únicos entre todas las fluctuaciones en la diferencia de potencial que pueden propagarse a lo largo de metros. Las fluctuaciones en la diferencia de potencial que están por debajo del umbral que hace disparar una neurona sólo se propagan distancias cercanas al milímetro, después

de haberse visto sometidas a atenuaciones. Por otro lado, los potenciales de acción son regenerados activamente, lo que permite que viajen grandes distancias sin apenas atenuación.

### 5.1.2. Frecuencia de disparo

A pesar de que los potenciales de acción tienen diferente forma, amplitud y duración, son tratados todos como eventos iguales. Si además ignoramos la corta duración de los spikes, se puede considerar que toda la información proveniente de los estímulos naturales es codificada en trenes de spikes. De tal forma que cada tren de spikes vendrá caracterizado por una lista de tiempos en los cuales ha habido un disparo. Matemáticamente estos trenes de spikes son expresados de la siguiente forma:

$$\rho(t) \equiv \sum_{i=1} \delta(t - t_i) . \quad (5.1.1)$$

A la función  $\rho(t)$  se le llama función de respuesta. Dado un mismo estímulo esta función de respuesta varía entre un registro y otro. Por ello que las respuestas neuronales suelen ser tratadas de forma estadística y por lo tanto vamos a necesitar alguna cantidad que contenga información sobre el promedio de todos los registros, y no sobre uno en concreto. Esta función se llama frecuencia de disparo,

$$\begin{aligned} r[t; s(\tau)] &\equiv \left\langle \sum_{i=1} \delta(t - t_i) \right\rangle = \\ &= \sum_N \int_0^T dt_1 \int_0^T dt_2 \cdots \int_0^T dt_N P[\{t_i\} | s(\tau)] \sum_{i=1}^N \delta(t - t_i) . \end{aligned} \quad (5.1.2)$$

### 5.1.3. Estimando la frecuencia de disparo

La actividad de una neurona en un instante  $t$  de tiempo, normalmente depende del comportamiento del estímulo durante un intervalo de tiempo, que empieza varios de cientos de milisegundos antes del instante  $t$  y termina una decena de milisegundos después de este instante. El problema básico es construir un estimador de  $r_{est}(t)$  de la

frecuencia de disparo  $r(t)$  que haya sido provocado por un estímulo  $s(t)$ . La manera más simple de construir este estimador es asumir que la frecuencia de disparo para cualquier instante de tiempo puede ser expresada como una suma ponderada de los valores que ha tomado el estímulo en tiempos pasados. Esta suma toma la forma de integral al considerar que la variable temporal es continua y se escribe

$$r_{est} = r_0 + \int_0^\infty d\tau D(\tau) s(t - \tau) . \quad (5.1.3)$$

El término  $r_0$  tiene en cuenta la actividad de fondo cuando no hay estímulo.  $D(\tau)$  es una función que cuantifica cuánto afecta el valor del estímulo a tiempo  $t - \tau$  a la frecuencia de disparo a tiempo  $t$ .

De una manera heurística hemos dado una justificación de los términos de la ecuación 5.1.3, aunque de manera más formal estos corresponden a los dos primeros términos de una expansión sistemática de la respuesta en potencias del estímulo. Esta expansión se le conoce con el nombre de expansión de Volterra, y para nuestro caso adopta la siguiente expresión

$$\begin{aligned} r_{est} = & r_0 + \int d\tau D(\tau) s(t - \tau) + \int d\tau_1 d\tau_2 D_2(\tau_1, \tau_2) s(t - \tau_1) s(t - \tau_2) + \\ & + \int d\tau_1 d\tau_2 d\tau_3 D_3(\tau_1, \tau_2, \tau_3) s(t - \tau_1) s(t - \tau_2) s(t - \tau_3) + \dots . \end{aligned} \quad (5.1.4)$$

Esta serie fue reordenada por Wiener [40] para poder calcular los términos de la serie de forma más sencilla. Los dos primeros términos de las expansiones de Volterra y Wiener son iguales, y por ello se conoce a  $D$  como el primer filtro de Wiener, el filtro lineal, o, cuando los términos de orden superior no son considerados, simplemente el filtro.

Para construir un estimador de la frecuencia de disparo basado sólo en los términos lineales, elegimos un filtro  $D$  que minimice la diferencia cuadrática media en un intervalo  $T$  entre el estimador y la frecuencia de disparo,

$$E = \frac{1}{T} \int_0^T dt (r_{est}(t) - r(t))^2 . \quad (5.1.5)$$

Esta minimización tiene como consecuencia que  $D$  debe satisfacer la siguiente relación

$$\int_0^\infty d\tau' Q_{ss}(\tau - \tau') D(\tau') = Q_{rs}(-\tau) , \quad (5.1.6)$$

siendo  $Q_{rs}(\tau)$  la función de correlación entre el estímulo y la frecuencia de disparo y  $Q_{ss}(\tau)$  la función de autocorrelación del estímulo. Este método que hemos descrito se conoce como *reverse correlation* porque la correlación entre la frecuencia de disparo y el estímulo es evaluada en  $-\tau$  en esta ecuación.

La ecuación 5.1.6 puede fácilmente resolverse si el estímulo que se utiliza está decorrelado, i.e

$$Q_{ss}(\tau) = \sigma_s^2 \delta(\tau) , \quad (5.1.7)$$

siendo  $\sigma_s^2$  la variancia del estímulo. A este tipo de estímulo se le conoce como ruido blanco.

Utilizando entonces este ruido blanco, se llega a que el filtro óptimo que proporciona la mejor estimación lineal de la frecuencia de disparo es

$$D(\tau) = \frac{Q_{rs}(-\tau)}{\sigma_s^2} . \quad (5.1.8)$$

Las neuronas y los sistemas nerviosos se comportan de manera no lineal, por tanto esta estimación lineal sólo es una aproximación. La predicción lineal tiene dos problemas: no hay nada que nos prevenga de predecir frecuencias de disparo negativas y no existe umbral de saturación para la frecuencia de disparo. Una manera de solucionar estos problemas es aplicar a esta estimación lineal una función no lineal

$$r_{est}(t) = r_0 + F(L(t)) , \quad (5.1.9)$$

donde  $L$  representa el término lineal

$$L(t) = \int_0^\infty d\tau D(\tau) s(t - \tau) . \quad (5.1.10)$$

## 5.2. Introducción al Sistema Visual Primario

La conversión de un estímulo luminoso en una señal eléctrica, que posteriormente provocará un potencial de acción, ocurre en la retina. En la figura 5.2(A.) se representa

el diagrama anatómico de la retina de un perro, donde se muestra también los cinco principales tipos de células retinianas. En la retina, la luz es convertida en una señal eléctrica por medio de los fotorreceptores, los conos y los bastones. En la figura 5.2(B.) se muestran registros intracelulares tomados en diferentes neuronas de la retina de una salamandra. El estímulo que se usó para los registros fue un pulso o flash luminoso de un segundo, que se hizo incidir contra la región de fotorreceptores de la izquierda en la figura 5.2(B.). Los bastones, especialmente el que está a la izquierda de la figura 5.2(B.), son hiperpolarizados por la luz. Esta señal eléctrica a través de las conexiones sinápticas es enviada hacia las células horizontales y bipolares. Uno puede darse cuenta como en una de estas células bipolares se invierte la tendencia hacia la despolarización. Estos cambios suaves en el potencial de membrana proporcionan una buena representación de la intensidad mientras se produce el estímulo. Esta forma de codificar la información es efectiva dentro de la retina, donde las distancias son pequeñas. Sin embargo no es adecuada cuando se trata de transmitir la señal de la retina al cerebro.

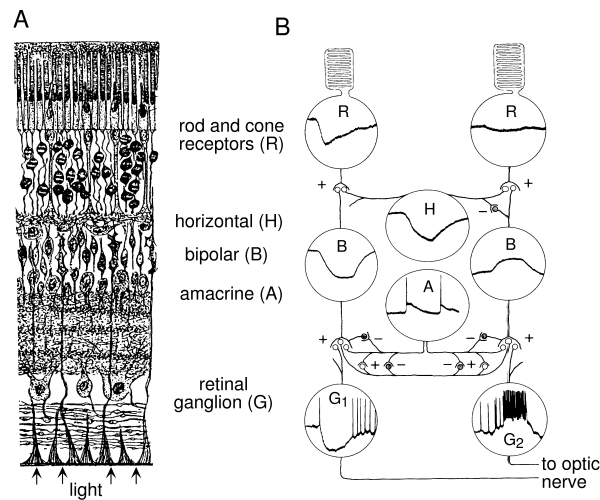


Figura 5.2: A. Diagrama del circuito de la retina de un perro. Los diferentes tipos de células se presentan a la derecha. Adaptado por [41] de unos dibujos de [42]. B. Registro intracelular de neuronas retinianas de una salamandra, respondiendo a un impulso de luz de un segundo [43].

Las células ganglionares van a ser las células que van a transmitir la información fuera de la retina a través del nervio óptico. Como vemos en la figura 5.2(B.), los registros de las células ganglionares son muy parecidos a los de las células bipolares, con la salvedad que en el caso de los registros de las células ganglionares hay también una transmisión de potenciales de acción. Las dos células ganglionares que se muestran en la figura presentan diferentes registros.  $G_2$  dispara cuando hay luz, mientras que  $G_1$  lo hace cuando no la hay. Es por ello que se denominan a las respuestas, ON y OFF. El nervio óptico transmite el tren de spikes de las células ganglionares al núcleo geniculado lateral (LGN), que actúa como una estación repetidora entre la retina y la corteza visual primaria, como se muestra en la figura 5.3. Antes de llegar al LGN, hay ciertas células ganglionares que se cruzan en el quiasma óptico. Esto permitirá que la parte izquierda y derecha de los campos visuales de los dos ojos sea representada en la parte izquierda y derecha del cerebro respectivamente.

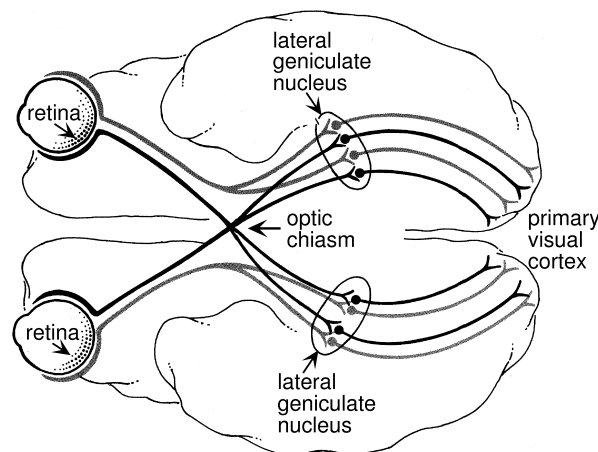


Figura 5.3: Vía sensorial desde la retina a la corteza visual primaria (V1), atravesando el núcleo geniculado lateral (LGN) del tálamo, en un cerebro humano [41].

### 5.3. Campos receptivos

El concepto tradicional de campo receptivo (*receptive field profile*) de una célula es una representación de la magnitud de respuesta de dicha célula a la presencia de un estímulo, en términos de un funcional sobre las coordenadas espaciales del estímulo, [36]. De tal forma, el campo receptivo (de una célula) es definido como una función (de las coordenadas espaciales) que cuantifica la actividad sináptica de esta célula bajo la influencia directa de un estímulo en una posición espacial determinada. En sentido estricto, definido de esta forma, sólo las células sensibles a los estímulos (las receptoras) serían susceptibles a tener campos receptivos (por su propia definición). Por ello, se generaliza el concepto: para una célula que no sea receptora el campo receptivo cuantifica la actividad sináptica de ésta con sus células receptoras de estímulos (aquellas con las que está conectada). En este último caso conviene hablar no ya de la posición espacial del estímulo sino de la situación de las células receptoras que son sensibles a dicho estímulo. Para mayor confusión, es usual primero definir al dominio (entendido como la región del espacio donde la función no se anula) del campo receptivo también con el mismo nombre: campo receptivo. En este trabajo cuando hablemos del dominio lo diremos explícitamente para una mejor comprensión del texto, reservando el nombre de campo receptivo a la función propiamente dicha.

#### 5.3.1. Campos receptivos visuales

Las neuronas en la retina, de LGN, y de la corteza visual primaria responden a estímulos luminosos en regiones restringidas del campo visual: el dominio del campo receptivo visual. Debido a que en el procesamiento sólo se tiene en cuenta las proyecciones polares de éstos estímulos, el dominio visual estará restringido a dos dimensiones polares. Cuando la luz estimula regiones que están fuera del dominio del campo receptivo de una neurona, ésta no producirá una respuesta directa, pero afectará, debido a las conexiones laterales, la respuesta de estímulos que estén dentro del dominio. La región que queda fuera del dominio del campo receptivo sólo desempeña un papel, inhibitorio o excitatorio, cuando el estímulo se encuentra también dentro del dominio. Para dife-

renciar las dos regiones -y crear todavía más confusión- se suele etiquetar al dominio como *campo receptivo clásico*, sin embargo, en este trabajo no se considerará la influencia de las regiones que no producen una respuesta directa: suprimimos las conexiones laterales.

Las dimensiones de estos dominios varían según qué neuronas estemos mirando, incluso dominios de campos receptivos asociados a neuronas de una misma clase, pueden tener tamaños diferentes dependiendo de la posición de estas neuronas. Así en el mono, el tamaño de los dominios asociados a células corticales varía entre unas décimas de grados cerca de la fovea a unos pocos grados en la periferia.

Para estimar la frecuencia de disparo de una neurona en respuesta a una imagen usaremos la ecuación 5.1.9. Debido a que el estímulo visual depende de la localización espacial, tenemos que decidir como contribuyen las diferentes posiciones de la imagen a la predicción lineal  $L(t)$ . La hipótesis más simple es asumir que las contribuciones de los distintos puntos espaciales se añaden de forma lineal, de manera que  $L(t)$  será obtenido integrando a todos los puntos del plano XY en el que se localiza la imagen

$$L(t) = \int_0^\infty \int \int dx dy d\tau D(x, y, \tau) s(x, y, t - \tau) . \quad (5.3.1)$$

Ahora, el filtro  $D(x, y, \tau)$  cuantifica cómo afecta el estímulo  $s$ , en el punto  $(x, y)$  y en un instante de tiempo  $t - \tau$ , a la frecuencia de disparo de la neurona en un instante de tiempo  $t$ . Este filtro *contiene* al campo receptivo visual de una neurona. Las células  $\mathcal{X}$  de la retina del gato y de LGN, células  $\mathcal{P}$  de la retina del mono y de LGN, así como las células simples de la corteza visual primaria satisfacen la hipótesis de linealidad. Sin embargo, las células  $\mathcal{Y}$  de la retina del gato y de LGN, células  $\mathcal{M}$  de la retina del mono y de LGN y células complejas de la corteza visual primaria necesitan incluir términos no lineales en la descripción. Hubel y Wiesel en su artículo [44] hicieron una clasificación similar: células simples son aquellas cuyos campos receptivos pueden ser descompuestos como suma de campos receptivos de células ON y OFF, las células complejas son aquellas cuyos campos receptivos no pueden ser descompuestos de esa manera.



### 5.3.2. Forma funcional de los campos receptivos visuales

El filtro  $D(x, y, \tau)$  es una función de tres variables; por ello es difícil de medir y de visualizar. Para ciertas neuronas, sin embargo, se puede considerar que este filtro se puede escribir como producto de dos funciones,

$$D(x, y, \tau) = D_s(x, y)D_t(\tau) , \quad (5.3.2)$$

una que describa la parte espacial: el campo receptivo (espacial) ,  $D_s(x, y)$ , y otra que describa la parte temporal,  $D_t(\tau)$ . Dado que sólo consideramos el procesamiento estático y no hay peligro de confusión no nos es necesario *apellidar* las dos partes.

Los células visuales van a ser selectivas a diferentes estímulos que provoquen fuertes respuestas. Las células glangionares de la retina y las neuronas de LGN: células ON (u OFF), tienen unos campos receptivos con simetría radial donde se distinguen dos zonas: una central que se activará (o inhibirá) cuando sea iluminada, rodeada de otra que actuará al contrario, se inhibirá (o activará) bajo la presencia de luz. Los estímulos *preferidos* serán, por tanto, puntos circulares de luz (o puntos oscuros rodeados de coronas circulares iluminadas). En la corteza visual primaria los campos receptivos son algo más complejos, como ahora vamos a ver, y suelen responder mejor a gradientes en el contraste -más o menos complejos- con una determinada orientación a la que las células serán selectivas. En la figura 5.4 podemos observar la selectividad a la orientación de una neurona de la corteza visual primaria de un mono: en el dominio del campo receptivo se presentaron barras de luz con diferentes orientaciones midiéndose la respuesta que provocaban.

La función de Gabor es una parametrización del campo receptivo de una célula simple de la corteza visual, [46] y [36]. Existen otras parametrizaciones para los campos receptivos; por ejemplo es frecuente utilizar la diferencia de gaussianas (DOG) para describir los campos receptivos de células ON y OFF de la retina y de LGN. La función de Gabor se construye multiplicando una gaussiana bidimensional por una senoide,

$$D_s(x, y) = A \exp \left( -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right) \cos(2\pi f x + \phi) , \quad (5.3.3)$$

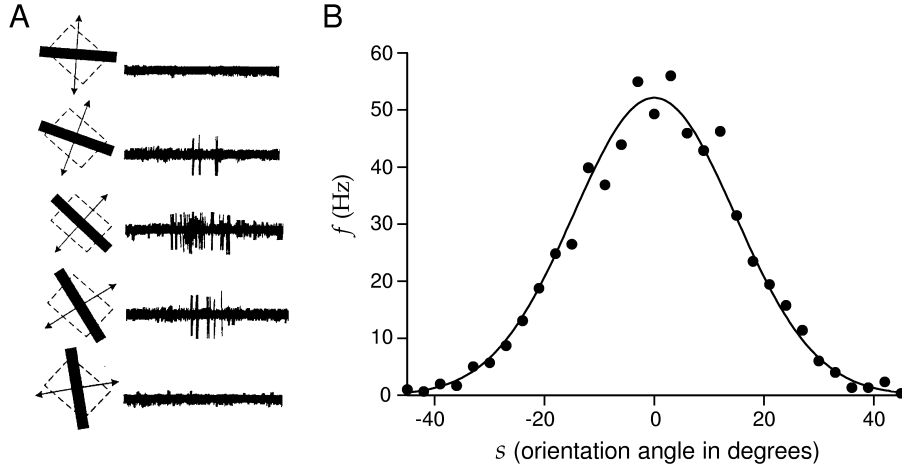


Figura 5.4: A. Actividades neuronales en la corteza visual primaria de un mono cuando una barra con determinadas orientaciones se mueve en el dominio del campo receptivo. B. Promedio de la frecuencia de disparo en una neurona de V1 de un gato en función de la orientación. (A. de [44] y B. de [45])

y tiene como parámetros:  $\sigma_x$  y  $\sigma_y$  representan las varianzas de las gaussianas de los ejes  $x$  e  $y$  respectivamente;  $f$  es la frecuencia espacial del coseno;  $\phi$  representa la fase espacial del coseno; y, por último,  $A$  es una constante de normalización. El sistema de coordenadas utilizado se ha elegido con un determinado centro y una determinada orientación. Para tener en cuenta todas las posibles traslaciones y rotaciones se introducen tres nuevos parámetros,

$$\begin{aligned} x' &= (x - x_0) \cos \theta + (y - y_0) \sin \theta \\ y' &= -(x - x_0) \sin \theta + (y - y_0) \cos \theta, \end{aligned} \quad (5.3.4)$$

donde  $\theta$  es el ángulo de la rotación y  $(x_0, y_0)$  es el punto de la traslación.

Para una discusión acerca de las limitaciones de la función de Gabor como aproximante matemático de los campos receptivos se invita al lector a que acuda al apéndice A.

En la figura 5.5 se representa los campos receptivos obtenidos en un experimento y la parametrización de estos campos receptivos usando la función de Gabor.

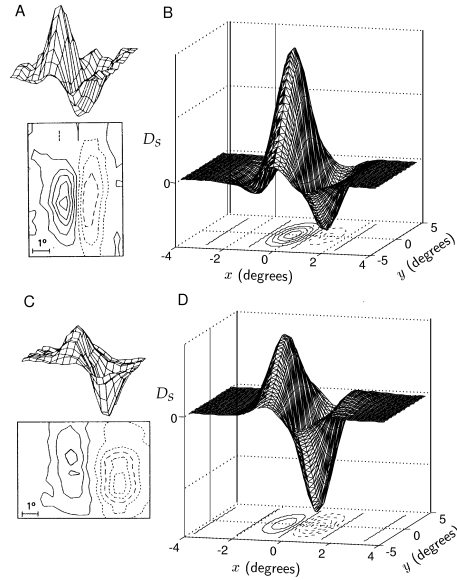


Figura 5.5: Estructura de los campos receptivos de células simples. A. y C. Estructura de los campos receptivos de dos neuronas de la corteza visual primaria de un gato, habiéndose determinado promediando el estímulo entre 50 ms y 100 ms antes del disparo del spike [47]. B. y D. Funciones Gabor con parámetros escogidos para representar las campos receptivos anteriores (figura extraída de [47]).

En la figura 5.6 se representa la distribución de los campos receptivos de células de V1 en el macaco, en unos ejes coordenados. Estos datos experimentales han sido medidos utilizando la técnica *subspace reverse correlation* [48], este método estima los campos receptivos en el espacio de Fourier, mientras que el *reverse correlation* lo hacía en espacio directo.

Los ejes coordenados del plano de la figura 5.6,  $n_x = \sigma_x f$  y  $n_y = \sigma_y f$ , permiten representar los datos experimentales a partir de la estimación de tres de sus parámetros, utilizando la función de Gabor,  $f$ ,  $\sigma_x$  y  $\sigma_y$ . Uno puede interpretar los valores  $n_x$  y  $n_y$  como el número de oscilaciones de la función de Gabor que hay en una longitud  $\sigma_x$  y  $\sigma_y$ , respectivamente. Con otras palabras, representan el tamaño de la gaussiana medido en unidades de periodo espacial,  $T = \frac{1}{f}$ .

Esta representación que utiliza Ringach es muy útil porque tiene una serie de simetrías. Es invariante respecto a cambios de escala de los campos receptivos. Si hacemos un cambio de escala con un factor  $\alpha$  tendremos que

$$n_x = (\alpha \sigma_x) \left( \frac{f}{\alpha} \right) = \sigma_x f . \quad (5.3.5)$$

Bajo traslaciones, rotaciones o cambios en la fase espacial de los campos receptivos, el plano  $(n_x, n_y)$  es invariante porque no depende de los parámetros  $x_0$ ,  $y_0$ ,  $\theta$  y  $\phi$

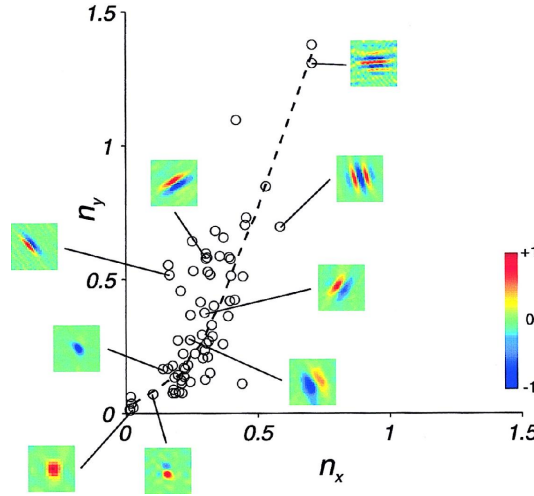


Figura 5.6: Distribución de la estructura espacial de los ajustes de los campos receptivos de células de V1 en el macaco (figura extraída de [39])

## 5.4. Descripción del modelo

En el capítulo 3 al clasificar las características visuales obtuvimos unas estructuras que evocaban a las funciones de Gabor. Eso nos hizo preguntarnos si eran susceptibles de ser las diferentes voces de una base de wavelets. Por otro lado, hemos revisado que las funciones Gabor sirven como función que parametriza los campos receptivos. El paso siguiente parece claro: ¿pueden nuestros filtros ser candidatos a describir de manera cualitativa los campos receptivos?.

Lo que proponemos no es *el* modelo que basado en propiedades estadísticas de las imágenes naturales ha adquirido el procesamiento visual para la codificación óptima de los estímulos visuales, sino investigar si *un* modelo que predice la existencia de un filtro óptimo en imágenes naturales es capaz de describir o reproducir de modo cualitativo algunas propiedades del sistema visual. La motivación viene del hecho de que el sistema visual debe de tener una estrategia -que codificará y compactificará la ingente cantidad de información- para poder adaptarse de manera eficiente el medio. En particular, uno se podría preguntar, por ejemplo, si el código deberá haber interiorizado que las imágenes son autosimilares y, con el fin de una optimización, habrá aprendido una fórmula que consiga eliminar la redundancia asociada a dicha propiedad estadística; y haya implementado el proceso multiplicativo. Sin embargo, la optimización que realiza el sistema visual viene limitada por sus propiedades anatómicas o funcionales que no se ha tenido en cuenta en la deducción de nuestro modelo.

La idea básica del modelo que proponemos es que los filtros o wavelets óptimos que hemos derivado del análisis de las imágenes naturales son capaces de extraer las componentes independientes de éstas: las variables  $\eta$  definidas como el cociente de las proyecciones de estos filtros (las variables  $\alpha$ ) a dos escalas consecutivas sólo serán independientes si tratamos con la base óptima. Además, hemos visto que estas wavelets son selectivos a diferentes patrones simples y orientaciones: las características visuales. Si nuestra representación de las imágenes naturales en función de las variables de  $\eta$  es el código más óptimo, ello impondrá un límite a la representación que utilice el sistema visual. Por ello, sabiendo que nuestro código es óptimo podemos identificar a los diferentes filtros deducidos como prototipos de campos receptivos en V1.

### 5.4.1. Simplificaciones

No nos preguntamos cómo el sistema visual ha adquirido su código óptimo. En principio, dejamos de un lado la dinámica de aprendizaje en una visión parecida a la de Attick, [6]: como una caja negra, sin explicar la génesis y evolución de la dinámica de la red neuronal que lo ha hecho posible. Tampoco tenemos en cuenta, a priori, las propiedades anatómicas o funcionales del sistema visual.

Hemos asumido que las imágenes naturales son los estímulos visuales, sin embargo, como ya ha quedado dicho anteriormente, éstas son sólo las proyecciones polares de las escenas reales. Por lo tanto, no estamos teniendo en cuenta la codificación que realiza el sistema visual de la profundidad. Así mismo, los efectos causados por la binocularidad como, por ejemplo, la selectividad de ciertas neuronas de V1 a estímulos visuales presentados en los dos ojos, [49], los consideraremos suficiente pequeños. La distancia focal usada en la obtención de las imágenes (28 mm) nos permitirá realizar tal hipótesis.

El colectivo de imágenes naturales que hemos utilizado pertenecientes a la base de datos de van Hateren [30] son imágenes en niveles de gris, restringiendo nuestro estudio a la información visual captada por un único fotoreceptor.

Nuestro modelo no puede predecir ninguna propiedad temporal al tratar con imágenes estáticas.

### 5.4.2. Prototipos de características visuales

La parametrización que usaremos para definir las distintas características visuales va a ser ligeramente modificada a la que se usó en el capítulo 3 pero se va a seguir basando en las mismas propiedades fenomenológicas observadas.

Las dos direcciones de las características visuales tienen claramente un comportamiento bien distinto: una dirección de máxima variación en el contraste que definirá la orientación de la característica visual -y, en esencia, a la propia característica visual-, y la otra, perpendicular a la primera, que la vaciamos de contenido y sólo la consideramos para poder definir nuestro modelo en  $2D$ . Esto es muy interesante, consideraremos que se puede simplificar las características visuales a su dirección no trivial y que sólo se tendrá en cuenta para calcular posteriormente el filtro asociado. Las funciones sigmoideas seguirán siendo la base de nuestra parametrización.

Para simplificar la notación usada para los diferentes prototipos definamos antes una serie de expresiones que nos serán útiles.

Primero, el decaimiento de las características visuales -en ambas direcciones- será parametrizado por una misma función gaussina de parámetros  $\sigma$  (dispersión) y  $x_0$  (lo-

calización). De esta forma, uniformizamos la continuidad de la característica visual en las dos direcciones:

$$\mathcal{G}_{[x_0;\sigma]}(x) \equiv e^{\frac{(x+x_0)^2}{2\sigma^2}}. \quad (5.4.1)$$

Además, necesitamos dos funciones auxiliares que *controlarán* la paridad de la característica visual.

Función impar:

$$\mathcal{E}_{[x_0;\beta;m]}^+(x) \equiv m^+ \tanh\left(\frac{\beta}{2}(x+x_0)\right), \quad (5.4.2)$$

siendo parámetro  $x_0$  el centro de simetría de la función y con  $\beta$  controlamos la pendiente de la única transición. Por último,  $m$  es una constante que toma el valor 0 para las características visuales impares y  $\frac{1}{2}$  para las pares.

Función par:

$$\mathcal{B}_{[x_0;\beta_i;\beta_j;d]}(x) \equiv \mathcal{E}_{[x_0+d/2;\beta_i;0.5]}^+(x)\mathcal{E}_{[x_0-d/2;\beta_j;0.5]}^-(x). \quad (5.4.3)$$

Podemos observar que la función par la hemos formado como producto de las dos anteriores tomando  $m = \frac{1}{2}$ . El parámetro  $x_0$  sigue siendo el centro de simetría,  $\beta_i$  y  $\beta_j$  parametriza las pendientes de las transiciones y la simetría de la función (valores muy distintos le alejan de la simetría par), y  $d$ , distancia entre los puntos de inflexión de las dos tangentes hiperbólicas, que da idea del tamaño de la parte central de la característica visual.

El prototipo más simple que podemos formar con los ingredientes introducidos es un borde simple (función impar más el decaimiento gaussiano); una única transición:

$$f_{[x_0;\beta;\sigma]}^1(x) \equiv \mathcal{E}_{[x_0;\beta;0]}^+(x)\mathcal{G}_{[x_0;\sigma]}(x), \quad (5.4.4)$$

y que sólo depende de 3 parámetros. Igualmente, si multiplicamos la función par por la gaussiana obtendremos la parametrización para las dos transiciones: una barra simple:

$$f_{[x_0;\beta_1;\beta_2;d;\sigma]}^2(x) \equiv \mathcal{B}_{[x_0;\beta_1;\beta_2;d]}(x)\mathcal{G}_{[x_0;\sigma]}(x), \quad (5.4.5)$$

que necesita dos parámetros más que el caso anterior. El perfil de tres transiciones -en principio impar- lo obtenemos multiplicando la anterior función por la función impar,

$$f_{[x_0;\beta_1;\beta_2;\beta_3;d;\sigma]}^3(x) \equiv \mathcal{E}_{[x_0;\beta_1;0]}^+(x) f_{[x_0;\beta_2;\beta_3;d;\sigma]}^2(x). \quad (5.4.6)$$

En esta ocasión, sólo necesitamos añadir un parámetro más. La función que parametriza las cuatro transiciones la definimos también a partir de  $f^2$  pero ahora multiplicada por la función par,

$$f_{[x_0;\beta_1;\beta_2;\beta_3;\beta_4;d_1;d_2;\sigma]}^4(x) \equiv \mathcal{B}_{[x_0;\beta_1;\beta_2;d_1]}(x) f_{[x_0;\beta_3;\beta_4;d_2;\sigma]}^2(x), \quad (5.4.7)$$

y otra vez volvemos a necesitar dos parámetros más. Siguiendo este procedimiento iterativo podemos construirnos el perfil para las cinco transiciones:

$$f_{[x_0;\beta_1;\beta_2;\beta_3;\beta_4;\beta_5;d_1;d_2;\sigma]}^5(x) \equiv \mathcal{E}_{[x_0;\beta_1;0]}^+(x) f_{[x_0;\beta_2;\beta_3;\beta_4;\beta_5;d_1;d_2;\sigma]}^4(x), \quad (5.4.8)$$

que, como en el caso  $f^3$ , necesitamos añadir un parámetro más respecto a la anterior parametrización.

Mediante esta forma de definición de los perfiles podemos generalizar a un número cualquiera de transiciones. Si teniendo definida la familia de funciones de  $2n$  transiciones estuviésemos interesados en construir la siguiente familia no tendríamos más que multiplicar la primera por la función impar,

$$f^{2n+1} \equiv \mathcal{E}^+ f^{2n}. \quad (5.4.9)$$

A su vez, la siguiente se definiría a partir de la función par,

$$f^{2n+2} \equiv \mathcal{B} f^{2n}. \quad (5.4.10)$$

En nuestro trabajo *sólo* hemos considerado las cinco primeras familias de las que mostramos un perfil concreto de cada una de ellas en la figura 5.7 a modo de ejemplo.

### 5.4.3. Filtros

Completaremos la parametrización de las características visuales proporcionando una continuación razonable a lo largo de la dirección perpendicular. Como se argumentó al principio de la anterior sección, escogeremos la gaussiana en la dirección  $y$  para extender las características visuales de la siguiente forma:



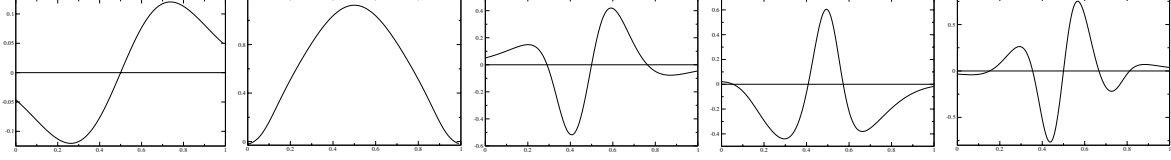


Figura 5.7: Cortes analíticos típicos del modelo. De izquierda a derecha:  $f^1$ ,  $f^2$ ,  $f^3$ ,  $f^4$  y  $f^5$ .

$$\langle c(x, y) \rangle \rfloor_r = f^r(x) \mathcal{G}_{[y_0; \sigma_y]}(y), \quad (5.4.11)$$

donde, como antes, los parámetros  $\sigma_y$  e  $y_0$  definen la dispersión y localización, respectivamente, de la gaussiana. Asignando un valor convencional unitario a la longitud de la característica visual en las dos direcciones: tenemos un tamaño de ventana efectiva de  $1 \times 1$  unidades cuadráticas; un valor impuesto para  $\sigma_x = 0.25$  y  $\sigma_y = 0.25$  (menor que el utilizado en el capítulo 3) es suficiente para garantizar que no haya solapes entre los filtros que se derivarán a partir de una pequeña modificación de la ecuación 4.4 para aplicarla a nuestras características visuales:

$$\Psi_r^\lambda(\vec{x}) = \left( \langle c(\vec{x}) \rangle \rfloor_r - \lambda \sum_{\vec{k} \in (\mathbb{Z}_2)^2} \langle c(2\vec{x} - \vec{k}) \rangle \rfloor_r \right). \quad (5.4.12)$$

Debido a nuestro desconocimiento de  $\lambda$ , debemos dejar, en principio, este parámetro libre, aunque por su propia definición éste está comprendido entre  $-0.5$  y  $0.5$ , nosotros sólo tomaremos tres valores paradigmáticos:  $\lambda = -0.5$  (antipersistencia total del signo),  $\lambda = 0$  y  $\lambda = 0.5$  (persistencia total del signo). De tal forma, que para cada familia de característica visual  $\langle c \rangle \rfloor_r$  tendremos todavía tres posibles filtros  $\Psi_r^\lambda$ .

#### 5.4.4. El efecto de LGN

Nuestra intención es comparar los filtros  $\Psi_r^\lambda$  con los campos receptivos medidos en V1. Sin embargo, entre V1 y la retina se produce un preprocesamiento que consiste en una decorrelación de las imágenes naturales: el efecto de LGN puede ser parametrizado

por un filtro de *whitening* proporcional a  $1/f$  (ver capítulo 2). Esta decorrelación es usada para eliminar redundancia y hacer el código más compacto como se discute en [6, 50].

Uno pensaría, entonces, que deberíamos decorrelar las imágenes antes de aplicar la ecuación 5.4.3. Sin embargo, como el filtro óptimo de imágenes naturales decorreladas se ve afectado de forma inversa, [29], es facil verificar que la proyección de la wavelet óptima de imágenes sin decorrelar sobre una imagen cualquiera,  $P_\Psi(I)$ , es proporcional a la proyección de la wavelet óptima de imagenes decorreladas sobre la misma imagen decorrelada,  $P_{\Psi^\omega}(I^\omega)$ . En la figura 5.8 podemos ver de manera esquemática lo discutido.

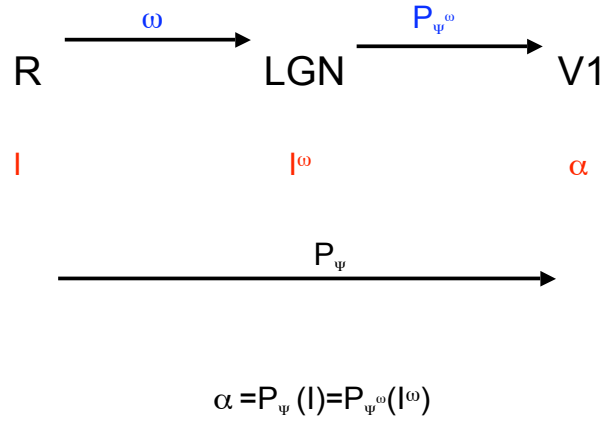


Figura 5.8: Esquema para explicar el efecto que tiene LGN en los campos receptivos de V1.

## 5.5. Resultados

Una vez que tenemos definido los filtros  $\Psi_r^\lambda$  a partir de las características visuales los ajustaremos a funciones Gabor definidas según la ecuación 5.3.3 y que identificaremos como  $\Gamma_r^\lambda$ . Como hemos visto, estas funciones vienen definidas por 8 parámetros. Tomando una serie de consideraciones podemos reducir a un ajuste de una función con sólo 5 parámetros:  $A$ ,  $\sigma_x$ ,  $\sigma_y$ ,  $f$  y  $\phi$ .

Los filtros que hemos definido los podemos orientar de tal forma que hagamos coincidir el eje del filtro que posee la tangente hiperbólica con el eje de la función Gabor definido por el coseno, de esta forma,  $\theta$  valdrá 0 y podremos eliminar un parámetro. Una vez que hemos tomado  $\theta = 0$  los parámetros  $x_0$  e  $y_0$  de nuestros filtros y de las funciones Gabor coinciden exactamente, así que no los tendremos en cuenta para el ajuste.

Para cada característica visual y los tres valores de  $\lambda$  se definieron una familia de 500 filtros obtenidos a partir de un muestreo aleatorio en los parámetros. Con el fin de conseguir una gran variedad de filtros y no dejar ningún grado de libertad optamos por que el muestreo en los parámetros fuese lo más extenso posible, abarcase hasta los valores extremos de los parámetros dentro de nuestro esquema  $([b_0, b_\infty], [d_0, d_{max}])$  y los filtros ya han saturado.

Así, los parámetros que modulan las pendientes de las tangentes hiperbólicas,  $\beta$ , toman valores comprendidos entre los extremos  $b_0 = 0.1$  y  $b_\infty = 30$ . Mientras, las distancias entre los puntos de inflexión varían entre  $d_0 = 0.05$  y una distancia máxima que viene dada por el tamaño de la ventana,  $d_{max} = 1$ .

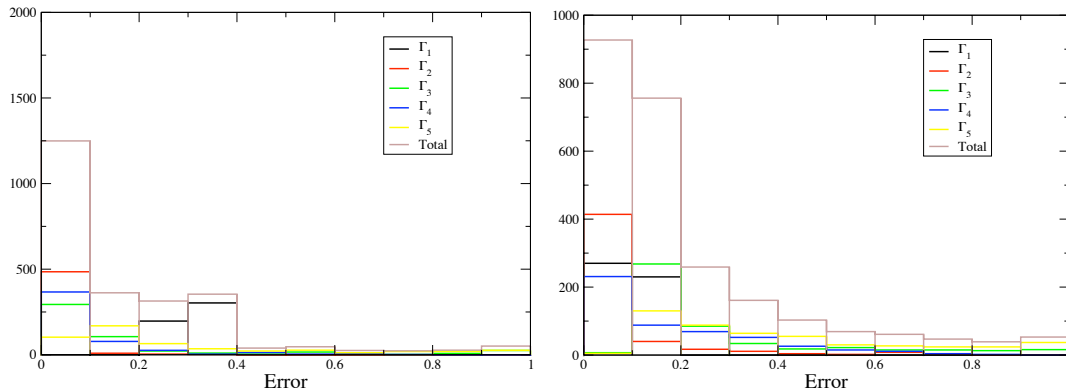


Figura 5.9: Izquierda. Histograma de los errores del ajuste de los campos receptivos según el modelo sigmoide con  $\lambda = 0.5$ . Derecha. Histograma de los errores del ajuste de los campos receptivos según el modelo sigmoide con  $\lambda = -0.5$ .

En las figuras 5.10 y 5.9 y representamos los histogramas de errores de los ajustes

de los 500 filtros por cada característica visual y su total para los tres valores de  $\lambda$ . Los errores vienen definidos como el cociente entre la desviación cuadrática del ajuste y la del filtro:  $\sigma_F^2/\sigma_\Psi^2$ . Se observa claramente que el mejor ajuste para las distintas características visuales se logra en el caso de  $\lambda = 0$ , aunque haya características visuales cuyos ajustes para otros valores de  $\lambda$  no sean mucho peores, como para  $\lambda = 0.5$  (persistencia total) los ajustes  $\Gamma_2$ ,  $\Gamma_3$  o  $\Gamma_4$ ; o para  $\lambda = -0.5$  (antipersistencia total) el ajuste  $\Gamma_1$ . Es por ello que nos fijaremos sólo en los ajustes obtenidos para  $\lambda = 0$  para hacer las distintas comparaciones con datos experimentales y con datos procedentes de otros modelos.

La distribución de los ajustes de campos receptivos experimentales de la figura 5.6 tienen unos errores (definidos de similar manera) cuyo histograma es bastante similar al obtenido por por nuestro modelo para  $\lambda = 0$  y que también mostramos en la figura 5.10. Aunque no podamos hacer ninguna comparación porque, primero, son pocos los puntos experimentales y, segundo, los puntos a ajustar son de naturaleza completamente distinta, nos sirve para imponer el mismo nivel de exigencia a nuestro modelo.

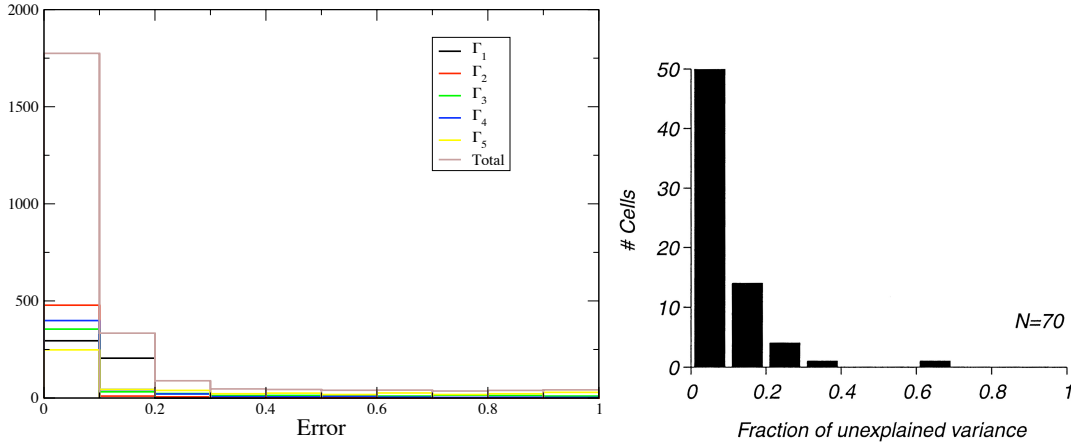


Figura 5.10: Izquierda. Histograma de los errores del ajuste de los campos receptivos según el modelo sigmoide con  $\lambda = 0$ . Derecha. Histograma de los errores de los ajustes de campos receptivos de células de V1 en el macaco (figura extraída de [39]).

### 5.5.1. Comparación con datos experimentales

Hemos representado la distribución de los filtros obtenidos con nuestro modelo en la figura 5.11. Para obtener cada punto del plano  $(n_x, n_y)$  hemos partido de los parámetros  $f$ ,  $\sigma_x$  y  $\sigma_y$  de la función de Gabor que mejor ajusta a cada filtro imitando el método que usó Ringach, [39]. Si comparamos nuestros resultados con los ajustes de datos experimentales de campos receptivos obtenidos por Ringach, figura 5.6, observamos que casi todos nuestros puntos también se alinean en rectas. No importa que las pendientes de estas rectas no sean las mismas, lo que hace que nuestro resultado sea bueno es que la distribución de los puntos no es aleatoria, sino que está restringida en una estrecha franja del plano  $(n_x, n_y)$ . Sin embargo, observamos que los puntos que están situados en la parte alta del plano tienen una dispersión mucho mayor.

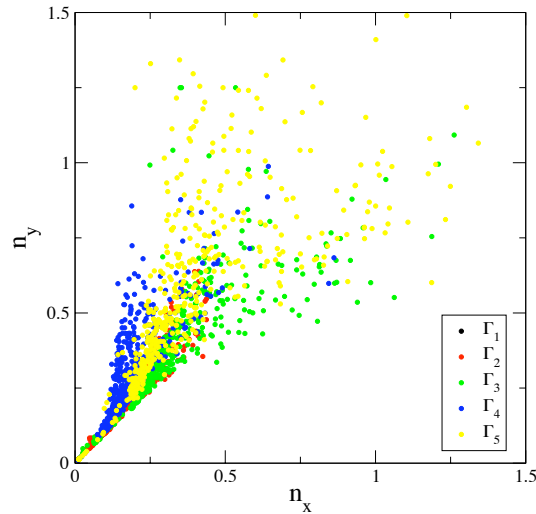


Figura 5.11: Distribución del ajuste de los campos receptivos en el plano  $(n_x, n_y)$  según el modelo sigmoide con  $\lambda = 0$ .

Para analizar la distribución de la fase espacial,  $\phi$ , parámetro al que no es sensible el plano  $(n_x, n_y)$ , haremos uso de las simetrías que tiene la función Gabor, [51], para definir una fase efectiva que toma como valor el argumento del numero complejo:  $|\cos \phi| + i|\sin \phi|$ . Esta nueva fase así definida tiene un rango de definición comprendido entre

0 y  $\pi/2$ . Podemos apreciar como nuestro modelo, figura izquierda 5.12, reproduce la bimodalidad presente en el experimento, figura derecha 5.12.

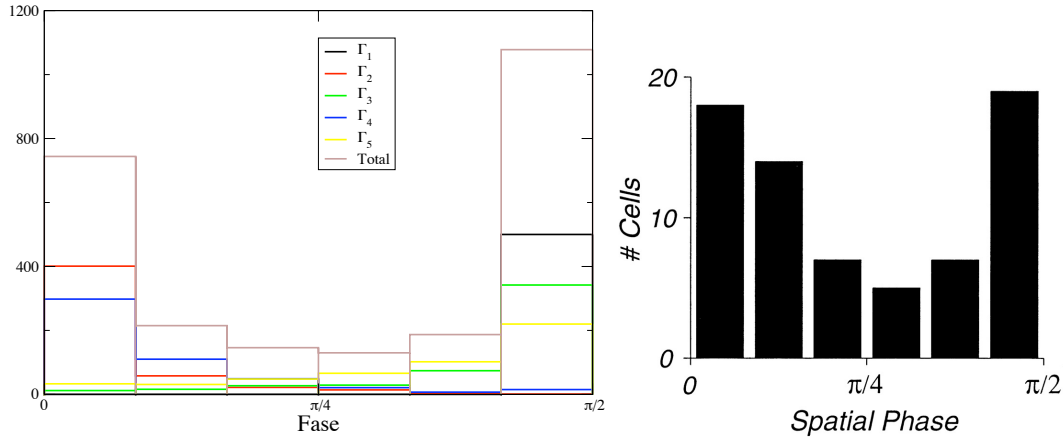


Figura 5.12: Izquierda. Histograma de las fases del ajuste de los campos receptivos según el modelo sigmoide con  $\lambda = 0$ . Derecha. Histograma de las fases de los campos receptivos de células de V1 en el macaco (figura extraída de [39]).

Aunque sólo hayamos comparado con unos registros experimentales que proceden de una especie determinado: un macaco. Experimentos realizados en la corteza visual del gato, [52] muestran que los puntos de los ajustes de los campos receptivos tienen un comportamiento extremadamente similar.

### 5.5.2. Comparación con otros modelos teóricos

Esto, sin embargo, no ocurre con otros modelos. En la figura 5.13 se representa la distribución de los campos receptivos de dos modelos distintos. Los puntos de estos modelos en el plano  $(n_x, n_y)$  están mucho más concentrados en una región, sobre todo en el caso del modelo de Análisis de Componentes Independientes.

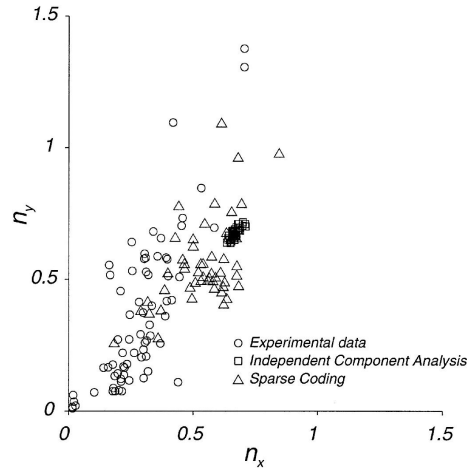


Figura 5.13: Comparación de la estructura de los campos receptivos entre datos experimentales [39] y dos modelos teóricos.

## 5.6. Conclusiones

Hemos visto que con un modelo simple basado en propiedades de las imágenes naturales conseguimos reproducir propiedades fundamentales de los campos receptivos de células simples en V1. Este modelo no posee ningún parámetro libre. La poca densidad que apreciamos en el plano  $(n_x, n_y)$  es susceptible de ser el lugar donde se situen filtros derivados de características visuales de más estructura.

Un estudio posterior deberá tener en cuenta la relevancia de la orientación de las características visuales en la predicción de los campos receptivos. La contribución de las diferentes características visuales (así como sus posibles orientaciones) al plano  $(n_x, n_y)$  ha de ser también investigado en un futuro trabajo.

No es un modelo que tenga en cuenta, *a priori*, las propiedades anatómicas o funcionales del sistema visual. Ni pretende afirmar que el procesamiento visual haya adquirido nuestro paradigma: el cómo procesa la información el sistema visual es una pregunta que no podemos contestar en el marco del presente trabajo. Sin embargo, los resultados evidencian que el código visual, por muy complejo que sea éste, ha adquirido ciertas simetrías estadísticas que tienen las imágenes naturales.

Las diferentes simplificaciones efectuadas: imágenes monocolors, monocularidad, ausencia de movimiento, etc, deberán ser consideradas para evaluar su importancia en un próximo estudio.



## Capítulo 6

# Resumen y Conclusiones

Las imágenes naturales son los estímulos del sistema visual y poseen ciertas simetrías cuyo conocimiento nos permite deducir propiedades del procesamiento. Para poder comprender cómo éste procesa y codifica la información procedente de estas imágenes, hemos analizado con detalle las propiedades de éstas. Mediante estas simetrías podemos definir un código eficiente por medio de una representación en wavelets que minimiza la redundancia. La base de wavelets la podemos extraer de una forma sencilla a través de las imágenes.

Los coeficientes de la base de wavelet no forman un código eficiente, pero debido a la multiescala en las imágenes podemos construir un código eficiente mediante el proceso multiplicativo que se deriva directamente de esta propiedad estadística.

La fórmula de la wavelet óptima funciona como un detector de características visuales en las imágenes naturales. Las imágenes están compuestas por distintas características visuales tales que según el colectivo que tomemos, predominarán unas u otras. Estas características visuales se pueden clasificar mediante una técnica sencilla, dando más peso a la idea de que en las imágenes existen estructuras que siguen una determinada jerarquía.

Los bordes y barras en las imágenes naturales son las características visuales que más predominan. Mediante un modelo simple hemos parametrizado estas características visuales. Esto nos ha permitido trabajar con un modelo fenomenológico, que aunque

muy simple, permite aglutinar en una función estas características visuales. Este modelo también soluciona de una manera eficaz el problema encontrado en la inferencia de información en las colas de la wavelet óptima. El siguiente paso a seguir debe ser una comprobación empírica del modelo sigmoide; sólo así tendremos la justificación de que nuestro modelo sirve para representar algunas propiedades de las imágenes.

Para mejorar los resultados obtenidos hemos revisado la fórmula de obtención del filtro óptimo: hemos comprendido mejor el problema, hemos formalizado las hipótesis y hemos propuesto una nueva formulación que, sin embargo, nos obliga a introducir un parámetro cuyo valor no podemos conocer *a priori*.

A pesar de las limitaciones, hemos comprobado como las imágenes naturales pueden ser descompuestas con filtros óptimos que se derivan de propiedades estadísticas de éstas a partir del concepto de característica visual, por otro lado estos filtros son susceptibles de ser ajustados por funciones Gabor dando unos resultados que están dentro de los datos experimentales. Con lo cual se ha verificado que existe un nexo entre los estímulos y el código.

En el trabajo futuro habrá que seguir por el mismo camino: con el modelo de medias signadas hallaremos el filtro para cada colectivo de imágenes naturales y lo usaremos para representar cada individuo de este colectivo así como ajustar estos filtros con Gabor para poder comparar con los datos experimentales.

# Apéndice A

## Descomposición de Señales Infinitas, Función Unidad y Métodos de Inferencia del Filtro Óptimo

### A.1. Desarrollo completo de una señal infinita

Sea  $c(\vec{x})$  una señal con media nula definida en todo el plano. Hagamos una descomposición multiresolución en funciones wavelet usando una base diádica,

$$c(\vec{x}) = \sum_{r=0}^{n-1} \sum_{j=-\infty}^{j=\infty} \sum_{(k_1, k_2) \in \mathcal{Z}} \alpha_{rjk} \psi_{rjk}(\vec{x}) , \quad (\text{A.1.1})$$

el índice  $j$  denota el nivel de escala; así, las escalas más pequeñas, las que darán cuenta de los detalles más finos, vendrán representadas por  $j$  positivos, y las escalas más grandes, que darán cuenta de los detalles más groseros, tendrán un  $j$  negativo. El vector  $\vec{k}$  denota las distintas posiciones adonde trasladaremos las wavelets. El índice  $r$ , que da cuenta del número de voces, es necesario al tratar con objetos bidimensionales.

La ecuación anterior la podemos separar en dos partes, una que contenga todas la

escalas con los detalles más groseros que una cierta escala (que denotamos  $A_0(\vec{x})$ ) que fijaremos por conveniencia con  $j = 0$ , y otra parte que incluya los detalles más finos (denotamos  $B_0(\vec{x})$ ),

$$c(\vec{x}) = A_0(\vec{x}) + B_0(\vec{x}) , \quad (\text{A.1.2})$$

con

$$A_0(\vec{x}) \equiv \sum_{r=0}^{n-1} \sum_{j=-\infty}^{j=-1} \sum_{(k_1, k_2 \in \mathbb{Z})} \alpha_{rj\vec{k}} \psi_{rj\vec{k}}(\vec{x}) , \quad (\text{A.1.3})$$

y

$$B_0(\vec{x}) \equiv \sum_{r=0}^{n-1} \sum_{j=0}^{j=\infty} \sum_{(k_1, k_2 \in \mathbb{Z})} \alpha_{rj\vec{k}} \psi_{rj\vec{k}}(\vec{x}) . \quad (\text{A.1.4})$$

La primera parte  $A_0(\vec{x})$  la podemos reescribir haciendo uso de las funciones unidad [27] ( $\mathcal{U}$ ), estas funciones son ortogonales a la base de las wavelets pero no tienen que tener media nula,

$$A_0(\vec{x}) = \sum_{(k_1, k_2 \in \mathbb{Z})} \beta_{-1\vec{k}} \mathcal{U}_{-1\vec{k}} . \quad (\text{A.1.5})$$

Lo interesante de estas funciones unidad es que aglutinan todas las escalas y voces, restringiendo entonces el sumatorio a las posiciones.

La segunda parte  $B_0(\vec{x})$  a su vez la podemos desglosar,

$$B_0(\vec{x}) = \sum_{r=0}^{n-1} \sum_{j=0}^{j=\infty} \sum_{\vec{k} \notin (\mathbb{Z}_{2^j})^2} \alpha_{rj\vec{k}} \phi_{rj\vec{k}}(\vec{x}) + \sum_{r=0}^{n-1} \sum_{j=0}^{j=\infty} \sum_{\vec{k} \in (\mathbb{Z}_{2^j})^2} \alpha_{rj\vec{k}} \psi_{rj\vec{k}}(\vec{x}) . \quad (\text{A.1.6})$$

la diferencia entre los dos sumatorios reside en que el segundo( que denotamos  $E_0(\vec{x})$ ) sólo contiene desplazamientos dentro del área central definida por la escala  $j = 0$  y el primer sumando (denotado  $D_0(\vec{x})$ ) el resto.

## A.2. Aproximaciones para llegar a la wavelet óptima

Fijémonos en  $c$  pero sólo en un entorno  $\Omega$  de su dominio; en este entorno habrá una escala con una wavelet asociada, cuyo tamaño característico coincide con su entorno  $\Omega$

(esta escala la fijaremos como  $j = 0$ ). Esto es, aunque las wavelets pueden ser funciones de soporte no compacto (no se anulan en ningún entorno) su rango de actuación es limitado al estar los valores grandes (en módulo) concentrados en una región finita. Teniendo en cuenta que las wavelets a escalas más grandes que la de partida tienen un comportamiento cada vez más constante en el entorno  $\Omega$ , y asumiendo ciertas condiciones de continuidad, podemos hacer una primera aproximación y considerar que  $A_0(\vec{x})$  actúa como una constante global, y que la podemos asumir dentro de la definición de señal. Por otra parte dado el decaimiento rápido de las wavelets podemos despreciar el primer sumatorio del segundo término de la ecuación A.1.6. Bajo estas aproximaciones podemos expandir  $c$  en un entorno  $\Omega$ ,

$$c(\vec{x}) = \tilde{A}_0 + E_0(\vec{x}) . \quad (\text{A.2.1})$$

Como ya hemos mencionado antes, redefinimos  $c$  para englobar la constante  $\tilde{A}_0$  y signamos.

Ahora asumiendo que  $\langle |\eta_r| \rangle = \langle |\eta| \rangle \forall r$  podemos invertir la ecuación A.2.1 para despejar  $\Psi$

$$\Psi(\vec{x}) = \mathcal{K} \frac{1}{\langle |\alpha_{0\vec{0}}| \rangle} \left( \langle E_0(\vec{x}) \rangle - \langle |\eta| \rangle \sum_{\vec{k} \in (\mathbb{Z}_2)^2} \langle E_0(2\vec{x} - \vec{k}) \rangle \right) , \quad (\text{A.2.2})$$

siendo  $\mathcal{K}$  una constante que se determinará al pedir la normalización de  $\Psi$ .

Esta ecuación que resulta es la formalización exacta de la que ya dedujimos, ecuación 2.5.4. La clave es que aquí hemos puesto el valor medio de  $E_0(\vec{x})$  que es una versión truncada de  $c$ . Sólo en el caso de que estemos en el entorno de definición  $\Omega$  de  $c$  las dos cantidades coincidirán. Sin embargo, como trabajamos con señales experimentales sólo conocemos un entorno finito de éstas, y para poder calcular  $\Psi$  necesitamos, debido al segundo sumando de la ecuación A.2.2, más información de la señal, puesto que nosotros no conocemos  $E_0$  si nos salimos del entorno  $\Omega$ . Veamos esta idea representada en la figura A.1

De manera que habrá que inferir de alguna forma la información que no tenemos sobre la señal, pero que necesitamos para poder obtener la wavelet óptima.



$$E_0(2\vec{x} - \vec{i})|_{\Omega}$$

$$E_0(2\vec{x} - \vec{j})|_{\Omega} \quad E_0(2\vec{x} - \vec{i} - \vec{j})|_{\Omega}$$

$$c(\vec{x}) = E_0(\vec{x})|_{\Omega} \quad \sum_{\vec{k} \in (\mathbb{Z}_2)^2} c(2\vec{x} - \vec{k}) \neq \sum_{\vec{k} \in (\mathbb{Z}_2)^2} E_0(2\vec{x} - \vec{k})|_{\Omega}$$

Figura A.1: Representación esquemática de la diferencia entre  $c$  y  $E_0$ . Sólo en el entorno  $\Omega$  ambas funciones coinciden.

### A.3. Métodos de inferencia

Después de haber hecho toda una serie de aproximaciones, todavía nos encontramos con un gran escollo a resolver: para obtener la wavelet óptima a partir de los datos de una señal necesitamos (debido al segundo sumando de la ecuación A.2.2) conocer más información de la señal que la que conocemos de los datos experimentales.

En el caso del espacio directo, ecuación 2.5.4, se hace la hipótesis de que la wavelet es de soporte compacto, por lo tanto no necesitamos más información de la señal de partida. Esta hipótesis es muy fácil de implementar, considera que fuera del entorno  $\Omega$  no hay nada, pero impone un condición bastante severa: la wavelet es de soporte compacto.

El método de Fourier, ecuación 2.5.4, sin embargo, sí que hace una estimación de la de la señal fuera del rango. En el caso de Fourier se interpola linealmente en el dominio frecuencial los datos ignotos a partir de los datos experimentales. Esto conduce a un decaimiento fuera del dominio que depende de las amplitudes de los armónicos sí representados, y que típicamente es inversamente proporcional a la distancia.





# Apéndice B

## Función Gabor: Limitaciones

En este apéndice nos proponemos hacer un estudio frecuencial de la función de Gabor, concretamente nos fijaremos sólo en el módulo de su transformada de Fourier, para hacer notar la existencia de diferentes regiones, que podrán tener implicaciones a la hora de ser utilizadas las funciones Gabor como aproximante matemático de los campos receptivos de células simples.

### B.1. Transformada de Fourier de la función Gabor

La función de Gabor en dos dimensiones adopta la siguiente forma

$$G(\vec{x}') = Ae^{-x'^2/2\Omega}e^{-y'^2/2\sigma}\cos(2\pi f_0x' + \varphi) . \quad (\text{B.1.1})$$

Observamos que depende de 5 parámetros independientes, a los que habrá que añadir 3 más que provienen de la definición de  $\vec{x}'$ ; éste se obtiene trasladando el sistema de coordenadas original por un vector constante  $\vec{x}_0$  y rotando éste un ángulo  $\theta$

$$\vec{x}' = \mathcal{R}_\theta(\vec{x} - \vec{x}_0) , \quad (\text{B.1.2})$$

donde  $\mathcal{R}_\theta$  el operador de rotación.

Para hacer la transformada de Fourier de la ecuación B.1.1 en la variable  $\vec{x}$  haremos uso del carácter lineal del operador  $\mathcal{R}_\theta$ , de tal forma que

$$\mathcal{F}_{\vec{x}}G(\mathcal{R}_\theta(\vec{x} - \vec{x}_0)) = e^{-i2\pi\vec{f}\cdot\vec{x}_0}\widehat{G}(\mathcal{R}_\theta(\vec{f})) , \quad (\text{B.1.3})$$

siendo  $\widehat{G}$  la transformada de Fourier de la ecuación B.1.1 en la variable  $\vec{x}'$ , que tiene la siguiente expresión matemática

$$\widehat{G}(\vec{f}) = A\pi\sqrt{\Omega\sigma}e^{-2\pi^2\sigma f_y^2}[e^{i\varphi}e^{-2\pi^2\Omega(f_x-f_0)^2} + e^{-i\varphi}e^{-2\pi^2\Omega(f_x+f_0)^2}] . \quad (\text{B.1.4})$$

## B.2. Módulo de la transformada de Fourier

El módulo de la anterior expresión adopta la siguiente forma

$$|\widehat{G}(\vec{f})| = A\pi\sqrt{2\Omega\sigma}e^{-2\pi^2[\sigma f_y^2 + \Omega(f_x^2 + f_0^2)]}\sqrt{\cosh(8\pi^2\Omega f_x f_0) + \cos 2\varphi} . \quad (\text{B.2.1})$$

Para expresar esta última función en términos de la variable que nos interesa,  $f_x$ , tendremos que hacer uso de la siguiente relación

$$|\mathcal{F}_{\vec{x}}G(\mathcal{R}_\theta(\vec{x} - \vec{x}_0))| = |\widehat{G}(\mathcal{R}_\theta(\vec{f}))| . \quad (\text{B.2.2})$$

Estas expresiones nos permiten concluir que el módulo sigue dependiendo de todos los parámetros. Bajo ciertos límites la ecuación B.2.1 se simplifica y tiene una dependencia en un número menor de parámetros. Así, por ejemplo, en el caso de que estemos con  $\Omega f_0 > 0.5$  esta función tomará la siguiente forma funcional

$$|\widehat{G}(\vec{f})| = A\pi\sqrt{\Omega\sigma}e^{-2\pi^2\sigma f_y^2}[e^{-2\pi^2\Omega(f_x-f_0)^2}] . \quad (\text{B.2.3})$$

Observamos, pues, que la fase  $\varphi$  ya no aparece y que, siempre en el límite en el que estamos,  $f_0$  se puede interpretar como el valor que maximiza la función.

## B.3. Diferentes regiones

La ecuación B.2.1 como función de  $f_x$  tiene un comportamiento que depende de dos de sus parámetros,  $f_0$  y  $\Omega$ . Esto se puede ver si analizamos como según qué valores tomen éstos, habrá un intervalo de valores de  $f_0$  donde la función podrá tener un máximo que aproximadamente coincida con el valor de  $f_0$  y un mínimo en  $f_x = 0$

(Región de alta frecuencia), otro intervalo donde existirá un máximo próximo a cero (sin coincidir con el valor de  $f_0$ ) y también un mínimo en  $f_x = 0$  (Región intermedia), o un tercer intervalo de valores que tome  $f_0$  donde sólo haya que un máximo en  $f_x = 0$  (Región de baja frecuencia).

Para el caso en que tomemos  $\Omega = 1$  y  $\varphi = 0$  se puede ver numericamente que si  $f_0 < 1/2\pi$  entonces no habrá mínimo. Expresado de otra forma: si hacemos la derivada -parcial de  $f_x$ - del módulo, e igualamos a cero obtendremos la siguiente expresión

$$f_x \cosh(8\pi^2 \Omega f_x f_0) + f_x \cos 2\varphi = f_0 \sinh(8\pi^2 \Omega f_x f_0) , \quad (\text{B.3.1})$$

esta ecuación siempre tiene solución para  $f_x = 0$ , pero para que haya otra solución es necesario que el producto  $\Omega f_0$  supere un cierto umbral, que en el caso  $\Omega = 1$  y  $\varphi = 0$  es cuando  $f_0 > 1/2\pi$ .

En la figura B.1 se puede observar graficamente lo comentado anteriormente.

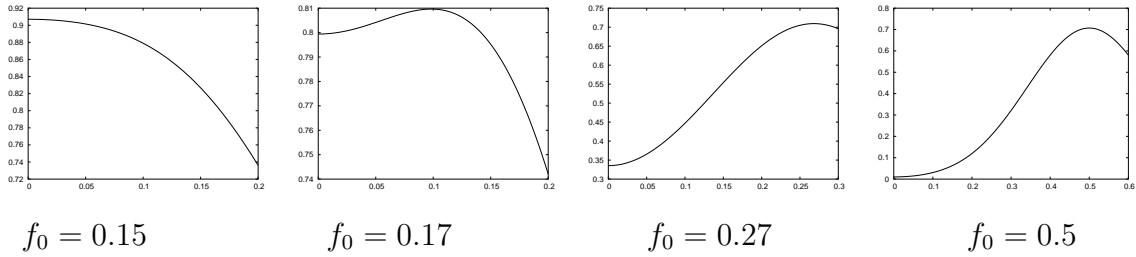


Figura B.1: Diferentes representaciones de la ecuación B.2.1 con  $A = 1/\pi$ ,  $\Omega = 1$ ,  $\sigma = 1/2$ ,  $\varphi = 0$  y  $f_y = 0$  como función de  $f_x$  para distintos valores de  $f_0$ .

## B.4. Estímulos y campos receptivos

Del análisis anterior se deducen dos importantes consecuencias, que tendrán relevancia a la hora de utilizar estas funciones Gabor como aproximante matemático del campo receptivo -espacial- de células simple.

La interpretación clásica que se da al parámetro  $f_0$  de la función de Gabor es de frecuencia preferida. Esto es, que un campo receptivo con una frecuencia preferida dada

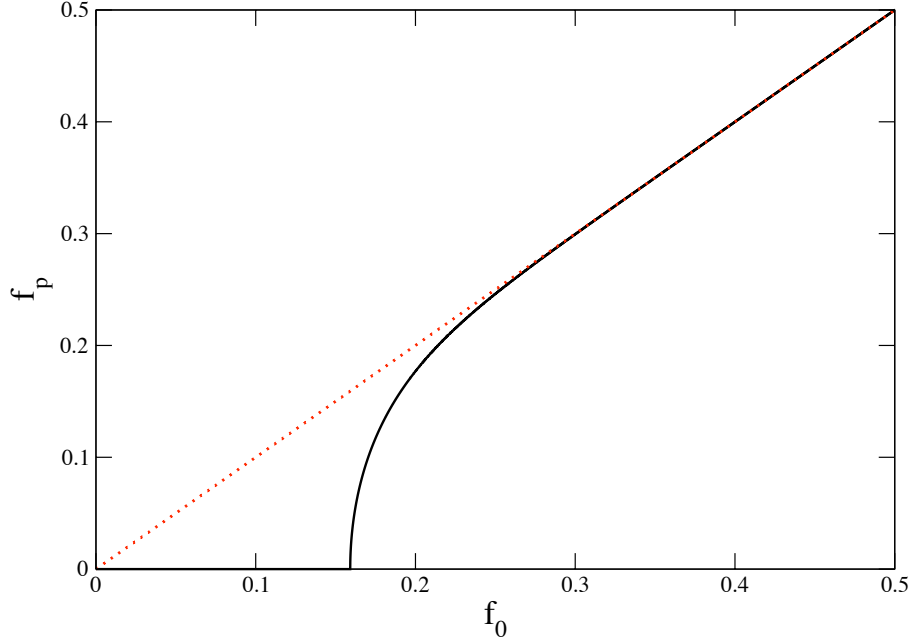


Figura B.2: Frecuencia preferida (aquella que maximiza la ecuación B.2.1) como función  $f_0$  con  $\varphi = 0$ .

(Función Gabor con un cierto parámetro  $f_0$ ) se le asocia un estímulo, rejilla sinusoidal que también tiene asociada una frecuencia  $f_p$ , con la siguiente forma

$$s(\vec{x}) = A_s \cos(2\pi f_p x - \phi) , \quad (\text{B.4.1})$$

que produce una respuesta máxima cuando justamente estas dos frecuencias coinciden,  $f_0 = f_p$ <sup>1</sup>. Sin embargo, esto sólo es cierto si nos situamos en la Región de alta frecuencia. En la Región intermedia el estímulo que maximiza la respuesta lineal tendrá una frecuencia menor que la frecuencia de la función Gabor. Si nos situamos en la Región de baja frecuencia para todas las funciones Gabor que pertenezcan a ese intervalo (recordemos, aquellas que estén por debajo de un cierto umbral) el único estímulo que maximiza la respuesta es el de frecuencia cero, es por tanto un estímulo constante espacialmente.

---

<sup>1</sup>No decimos nada de los demás parámetros por centrar estas notas en un análisis frecuencial

Si uno quiere ajustar datos experimentales con un modelo que viene definido a partir de varios parámetros, uno espera que éste sea sensible a las variaciones de los distintos parámetros. De esta forma, los datos experimentales podrán ser ajustados de forma casi unívoca por un modelo con unos parámetros dados. Sin embargo, hemos visto que en la Región de baja frecuencia el parámetro  $f_0$  pierde el valor *informativo* que antes poseía, y con ello la validez para ser un buen parámetro de ajuste: si uno se mueve en todo el intervalo la forma funcional (el máximo siempre va a estar en el mismo sitio) apenas va a sufrir cambios. Además, será bastante complicado extraer a partir de los datos experimentales un buen criterio para poder calcular que frecuencia,  $f_0$ , es la que se le puede asociar a dichos datos, siempre que nos encontremos en esta Región de baja frecuencia.



# Bibliografía

- [1] H. B. Barlow, “Possible principles underlying the transformation of sensory messages,” in *Sensory Communication*, W. Rosenblith, Ed., p. 217. M.I.T. Press, Cambridge MA, 1961.
- [2] F. Attneave, “Some informational aspects of visual perception,” *Psychological Review*, vol. 61, pp. 183–193, 1954.
- [3] R. Linsker, “Self-organization in a perceptual network,” *Computer*, vol. 21, pp. 105, 1988.
- [4] J.-P. Nadal, N. Brunel, and N. Parga, “Nonlinear feedforward networks with stochastic outputs: infomax implies redundancy reduction,” *Network: Computation in Neural Systems*, vol. 9, pp. 1–11, 1998.
- [5] S. B. Laughlin, “A simple coding procedure enhances a neuron’s information capacity,” *Z. Naturf.*, vol. 36, pp. 910–912, 1981.
- [6] J. J. Atick, “Could information theory provide an ecological theory of sensory processing?,” *Network: Comput. Neural Syst.*, vol. 3, pp. 213–251, 1992.
- [7] J. H. van Hateren, “Theoretical predictions of spatiotemporal receptive fields of fly lms, and experimental validation,” *J. Comp. Physiology A*, vol. 171, pp. 157–170, 1992.
- [8] D. J. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *J. Opt. Soc. Am.*, vol. 4, pp. 2379–2394, 1987.

- [9] F. Rieke, D. A. Bodnar, and W. Bialek, "Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents," in *Proc. R. Soc.*, 1995, vol. 262 of *London B*, pp. 259–265.
- [10] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.
- [11] A. J. Bell and T. J. Sejnowski, "The independent components of natural scenes are edge filters," *Vision Research*, vol. 37, pp. 3327–3338, 1997.
- [12] E. P. Simoncelli and O. Schwartz, "Image statistics and cortical normalization models," in *Advances in Neural Information Processing Systems*, 1999, vol. 11 of *MS Kearns, SA Solla, DA Cohn*, pp. 153–159.
- [13] A. Turiel, J.-P. Nadal, and N. Parga, "Orientational minimal redundancy wavelets: from edge detection to perception," *Vision Research*, vol. 43, no. 9, pp. 1061–1079, 2003.
- [14] J.-P. Nadal and N. Parga, "Duality between learning machines: a bridge between supervised and unsupervised learning," *Neural Computation*, vol. 6, pp. 491–508, 1994.
- [15] A Bell and T Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.
- [16] B.A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vision Research*, vol. 37, pp. 3311–3325, 1997.
- [17] D. L. Ruderman, "The statistics of natural images," *Network*, vol. 5, pp. 517–548, 1994.
- [18] D. J. Tollhurst, Y. Tadmor, and Ch. Tang, "Amplitude spectra of natural images," *Ophthal. Physiol. Opt*, vol. 12, pp. 229–232, 1992.



- [19] D. L. Ruderman and W. Bialek, “Statistics of natural images: Scaling in the woods,” *Physical Review Letters*, vol. 73, pp. 814, 1994.
- [20] A. Turiel, G. Mato, N. Parga, and J. P. Nadal, “Self-similarity properties of natural images,” in *Proc. of NIPS’97*. 1997, vol. 10, pp. 836–842, MIT Press.
- [21] A. Turiel, G. Mato, N. Parga, and J. P. Nadal, “The self-similarity properties of natural images resemble those of turbulent flows,” *Physical Review Letters*, vol. 80, pp. 1098–1101, 1998.
- [22] A. Turiel and N. Parga, “The multi-fractal structure of contrast changes in natural images: from sharp edges to textures,” *Neural Computation*, vol. 12, pp. 763–793, 2000.
- [23] A. Nevado, A. Turiel, and N. Parga, “Scene dependence of the non-gaussian scaling properties of natural images,” *Network*, vol. 11, pp. 131–152, 2000.
- [24] E. A. Novikov, “Infinitely divisible distributions in turbulence,” *Physical Review E*, vol. 50, pp. R3303, 1994.
- [25] B. Castaing, “The temperature of turbulent flows,” *Journal de Physique II*, vol. 6, pp. 105–114, 1996.
- [26] Z. Li and J. J. Atick, “Towards a theory of the striate cortex,” *Neural Computation*, vol. 6, pp. 127–146, 1994.
- [27] I. Daubechies, *Ten lectures on wavelets*, CBMS-NSF Series in Ap. Math. Capital City Press, Montpelier, Vermont, 1992.
- [28] Stéphane Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd Edition, 1999.
- [29] A. Turiel and N. Parga, “Multifractal wavelet filter of natural images,” *Physical Review Letters*, vol. 85, pp. 3325–3328, 2000.

- [30] J. H. van Hateren and A. van der Schaaf, “Independent component filters of natural images compared with simple cells in primary visual cortex,” *Proc. R. Soc. Lond.*, vol. B265, pp. 359–366, 1998.
- [31] A. Turiel, J.M. D. Delgado, and N. Parga, “Learning efficient internal representations from natural image collections,” *Neurocomputing*, vol. 58-60, pp. 915–921, 2004.
- [32] B. Dubrulle, “Intermittency in fully developed turbulence: Log-poisson statistics and generalized scale covariance,” *Physical Review Letters*, vol. 73, pp. 959–962, 1994.
- [33] L. Kaufman and P.J. Rousseeuw, *Finding groups in data: An introduction to cluster analysis*, John Wiley and Sons, 1990, ISBN 0471878766.
- [34] J.M. D. Delgado, A. Turiel, and N. Parga, “Receptive fields of simple cells from a taxonomic study of natural images and suppression of scale redundancy,” *Neurocomputing*, vol. 69, pp. 1224–1227, 2006.
- [35] J.M. D. Delgado, A. Turiel, and N. Parga, “Visual features and local optimal wavelet filters: A taxonomic study on a large image databasis,” En preparación, 2007.
- [36] J. G. Daugman, “Two-dimensional spectral analysis of cortical receptive field profile,” *Vision Research*, vol. 20, pp. 847–856, 1980.
- [37] H. B. Barlow, “Sensory mechanisms, the reduction of redundancy, and intelligence,” in *Mechanisation of Thought Processes*, vol. 2, pp. 537–574. Her Majesty’s Stationary Office, London, 1958.
- [38] A. Srivastava, A. Lee, E. Simoncelli, and S.-C. Zhu, “On advances in statistical modeling of natural images,” *J. of Mathematical Imaging and Vision*, vol. 18, pp. 17–33, 2003.

- [39] D. L. Ringach, “Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex,” *J. Neurophysiology*, vol. 2002, pp. 455–463, 2002.
- [40] N. Wiener, *Nonlinear Problems in Random Theory*, Wiley., New York, 1958.
- [41] J. G. Nicholls, R. Martin, and B. G. Wallace, *From Neuron to Brain: A Cellular and Molecular Approach to the Function of the Nervous System*, Sinauer Associates., Sunderland, MA, 1992.
- [42] S. Ramón y Cajal, *Histologie du Système Nerveux de l’Homme et des Vertébrés*, Maloine., Paris, 1911.
- [43] J. E. Dowling, *An Introduction to Neuroscience*, Bellknap Press., Cambridge, MA, 1992.
- [44] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” *J. Physiol.*, vol. 195, pp. 215–244, 1968.
- [45] G. H. Henry, B. Dreher, and P. O. Bishop, “Orientation specificity of cells in cat striate cortex,” *Journal of Neurophysiology*, vol. 37, pp. 1394–1409, 1974.
- [46] S. Marcelja, “Mathematical description of the responses of simple cortical cells,” *J. Optical Soc. Am.*, vol. 70, pp. 1297–1300, 1980.
- [47] J. Jones and L. Palmer, “The two-dimensional spatial structure of simple receptive fields in cat striate cortex,” *Journal of Neurophysiology*, vol. 58, pp. 1187–1211, 1987.
- [48] D. L. Ringach, G. Sapiro, and R. Shapley, “A subspace reverse correlation method for the study of visual neurons,” *Vision Research*, vol. 37, pp. 2455–2464, 1997.
- [49] G. F. Poggio and W. H. Talbot, “Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey,” *Journal of Physiology*, vol. 315, pp. 469–492, 1981.

- [50] Dan Yang, Joseph J. Atick, and R. Clay Reid, “Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory,” *J. of Neuroscience*, vol. 16, no. 10, pp. 3351–3362, 1996.
- [51] D. J. Field and D. J. Tolhurst, “The structure and symmetry of simple-cell receptive-field profiles in the cat’s visual cortex,” *Proc R Soc Lond B Biol Sci*, vol. 228, pp. 379–400, 1986.
- [52] J. Jones and L. Palmer, “An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex,” *Journal of Neurophysiology*, vol. 58, pp. 1233–1258, 1987.

# Índice de figuras

2.1. Curva de Koch. Ejemplo de un objeto autosimilar afín. . . . .	20
2.2. Triángulo de Sierpinski, objeto autosimilar afín. Cada uno de los tres triángulos con distintos colores que forman el triángulo de Sierpinski son iguales entre sí e iguales a este último. . . . .	21
2.3. Imagen de Lena . . . . .	23
2.4. Imagen de Lena tras multiplicar por $ \vec{f} $ en el espacio de Fourier (aproximador de decorrelación). . . . .	24
2.5. Imagen de Lena, habiendo eliminado la estadística de orden superior a dos. . . . .	26
2.6. Imagen de Lena, eliminado la estadística de orden dos. . . . .	27
2.7. Construcción de la wavelet individual a partir de la imagen <i>imk00040.imc</i> de la base de datos de Hans van Hateren (ver [30] para detalles sobre la imagen) . . . . .	35
2.8. Construcción de la wavelet a partir de un colectivo de 4167 imágenes de la base de datos de Hans van Hateren [30] . . . . .	36
2.9. Wavelet obtenida a partir de la fórmula 2.5 sobre un colectivo de 4167 imágenes de la base de datos de Hans van Hateren [30]. . . . .	37
2.10. Wavelet obtenida a partir de la figura 2.9 implementando sus simetrías. . . . .	38
2.11. Cortes, horizontal y vertical, de la wavelet de la figura 2.10 . . . . .	38

- 2.12. Verificación experimental del proceso multiplicativo, (2.4.5); comparación entre dos wavelets. **Izquierda:** Wavelet de Haar; **Derecha:** Wavelet óptima obtenida a partir de las 4167 imágenes, (2.10). Para el estudio se han utilizado un colectivo de 10 imágenes de la base de van Hateren. Empezando por el histograma (+) de los coeficientes de wavelet  $\alpha_{j,k}$  a la escala  $j = 5$ , asumiendo invariancia de traslación, y usando una distribución log-Poisson [21] de parametros:  $\Delta = 0.33$ ,  $\beta = 0.66$  y  $s = 1$  [22], obtenemos una predicción para la distribución de los  $\alpha_{j,k}$  a la escala  $j = 6$  (línea punteada). Deberemos compararlo con la obtención directa de la distribución de los  $\alpha_{j,k}$  a la escala  $j = 6$  (x). . . . . 39
- 2.13. Imagen *imk00040.imc* y tres representaciones de ésta utilizando los coeficientes  $\alpha$  a tres escalas diferentes. . . . . 41
- 2.14. Imagen *imk03236.imc* y tres representaciones de los los coeficientes  $\eta$  a tres escalas diferentes. . . . . 42
- 2.15.  $\sum_{j\vec{k}} \alpha_{j\vec{k}} \Psi_{j\vec{k}}(\vec{x})$  para  $j = 0, j \leq 2, j \leq 6$  y  $j \leq 8$  para las imágenes *imk03236.imc* y de Lena . . . . . 43
- 2.16.  $\sum_{rj\vec{k}} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x})$  para  $j = 0, j \leq 2, j \leq 6$  y  $j \leq 8$  para las imágenes *imk03236.imc* y de Lena con 2 orientaciones . . . . . 47
- 2.17.  $\sum_{rj\vec{k}} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x})$  para  $j = 0, j \leq 2, j \leq 6$  y  $j \leq 8$  para las imágenes *imk03236.imc* y de Lena con 3 orientaciones . . . . . 48
- 3.1. Fracción de vectores ( $\Lambda$ ) en función del solape ( $Q$ ) para espacios vectoriales de distintas dimensiones (2, 3, 4, 5, 6 y 256). . . . . 59
- 3.2. Imagen *imk00017.imc* de la base de datos de Hans van Hateren (ver [30] para detalles sobre la imagen). . . . . 60
- 3.3. Mapas con los elementos  $16 \times 16$  seleccionados para cada una de las cinco clases más representadas de la imagen *imk00017.imc* con  $Q = 0.3$  (**arriba**) y con  $Q = 0.7$  (**abajo**). Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. 61

- 3.4. **Arriba:** Imagen-colectivo *imk00005.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 63
- 3.5. **Arriba:** Imagen-colectivo *imk00016.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 64
- 3.6. **Arriba:** Imagen-colectivo *imk00017.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 65
- 3.7. **Arriba:** Imagen-colectivo *imk00019.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 66
- 3.8. **Arriba:** Imagen-colectivo *imk00207.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 67
- 3.9. **Arriba:** Imagen-colectivo *imk00406.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 68

- 3.10. **Arriba:** Imagen-colectivo *imk00408.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 69
- 3.11. **Arriba:** Imagen-colectivo *imk00802.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 70
- 3.12. **Arriba:** Imagen-colectivo *imk03813.imc*. **Medio:** Los mapas de cada clase con las subimágenes escogidas. Blanco: bloques no incluidos en la clase. Negro: bloques de la clase que están acoplados positivamente con la característica visual. Gris: bloques de la clase negativamente acoplados. **Abajo:** Características visuales asociados a cada clase de la línea central. 71
- 3.13. Imagen *imk00802.imc* de la base de datos de Hans van Hateren (ver [30] para detalles sobre la imagen). . . . . 72
- 3.14. Imágenes de las representaciones (**abajo**) de la imagen *imk00802.imc* usando los filtros (**arriba**) derivados de las características visuales de las cinco clases más representativas de esa misma imagen. Los PSNR individuales son 23.08 dB, 21.66 dB, 21.16 dB, 21.58 dB y 21.23 dB, respectivamente. . . . . 73
- 3.15. Representación usando el banco entero de filtros. El PSNR para esta representación es de 24.27 dB. . . . . 74
- 3.16. **Izquierda:** Característica visual de la primera clase de la imagen *imk00017.imc*. **Centro:** Wavelet obtenida aplicando la fórmula en el espacio de directo. **Derecha:** Wavelet obtenida aplicando la fórmula en espacio de Fourier. 75
- 3.17. Imagen *imk00005.imc* (izquierda), característica visual asociada a la primera clase (centro) y su corte vertical (derecha). . . . . 75



- 3.18. Imagen *imk00019.imc* (izquierda), característica visual asociada a la segunda clase (centro) y su corte vertical (derecha). . . . . 76
- 3.19. Imagen *imk00802.imc* (izquierda), característica visual asociada a la primera clase (centro) y su corte horizontal (derecha). . . . . 76
- 3.20. Imagen *imk00802.imc* (izquierda), característica visual asociada a la segunda clase (centro) y su corte horizontal (derecha). . . . . 77
- 3.21. Cortes analíticos típicos del modelo.  $f^0$  con  $\beta = 2$ ,  $\beta_\infty = 4$  (**izquierda**),  $f^1$  con  $\beta = 2$ ,  $x_0 = 0.1$  (**centro**) y  $f^2$  con  $\beta = 1$ ,  $x_0 = 0.2$ ,  $\beta_\infty = 4$ ,  $x_\infty = 0.2$  (**derecha**). . . . . 77
- 3.22. **Arriba:**  $R_{opt}$  para varios niveles de resolución. De izquierda a derecha:  $j = 0$  ( $PSNR = 19.70$  dB),  $j \leq 2$  ( $PSNR = 20.50$  dB) y  $j \leq 4$  ( $PSNR = 20.88$  dB). **Abajo:**  $R_{opt}$  para los niveles  $j \leq 6$  (izquierda,  $PSNR = 21.75$  dB),  $j \leq 8$  (centro,  $PSNR = 24.63$  dB) e imagen original *imk00802.imc*. . . . . 79
- 3.23. **Arriba:**  $R_{opt}$  para varios niveles de resolución. De izquierda a derecha:  $j = 0$  ( $PSNR = 23.02$  dB),  $j \leq 2$  ( $PSNR = 23.74$  dB) y  $j \leq 4$  ( $PSNR = 25.03$  dB). **Abajo:**  $R_{opt}$  para los niveles  $j \leq 6$  (izquierda,  $PSNR = 27.61$  dB),  $j \leq 8$  (centro,  $PSNR = 31.51$  dB) e imagen original *imk01409.imc*. . . . . 80
- 3.24. Todas las características visuales de las 500 imágenes seleccionadas: esta imagen contiene la información acerca de 2500 características visuales. Cada imagen da lugar a 5 características visuales, que se colocan formando una columna vertical. El nivel de intensidad del brillo de cada pixel es proporcional a la respuesta de la característica visual asociada, brillante para valores positivos y oscuro para negativos. El color sobreimpuesto es una indicación del número de subimágenes en la clase, relativo al total de 6144 subimágenes por colectivo (imagen). Se ha usado el siguiente código de colores: 0 a 1 %: Ocre; 1 a 5 %: azul; 5 a 10 %: rojo; 10 a 20 %: magenta; 20 a 30 %: verde; 30 a 40 %: cyan; 40 a 50 %: amarillo; más del 50 %: blanco. . . . . 84

5.1.	A. Registro intracelular de un potencial de acción de una neurona cortical de una rata. B. Diagrama de una sinapsis. . . . .	101
5.2.	A. Diagrama del circuito de la retina de un perro. Los diferentes tipos de células se presentan a la derecha. Adaptado por [41] de unos dibujos de [42]. B. Registro intracelular de neuronas retinianas de una salamandra, respondiendo a un impulso de luz de un segundo [43]. . . . .	105
5.3.	Vía sensorial desde la retina a la corteza visual primaria (V1), atravesando el núcleo geniculado lateral (LGN) del tálamo, en un cerebro humano [41]. . . . .	106
5.4.	A. Actividades neuronales en la corteza visual primaria de un mono cuando una barra con determinadas orientaciones se mueve en el dominio del campo receptivo. B. Promedio de la frecuencia de disparo en una neurona de V1 de un gato en función de la orientación. (A. de [44] y B. de [45]) . . . . .	110
5.5.	Estructura de los campos receptivos de células simples. A. y C. Estructura de los campos receptivos de dos neuronas de la corteza visual primaria de un gato, habiéndose determinado promediando el estímulo entre 50 ms y 100 ms antes del disparo del spike [47]. B. y D. Funciones Gabor con parámetros escogidos para representar las campos receptivos anteriores (figura extraída de [47]). . . . .	111
5.6.	Distribución de la estructura espacial de los ajustes de los campos receptivos de células de V1 en el macaco (figura extraída de [39]) . . . . .	112
5.7.	Cortes analíticos típicos del modelo. De izquierda a derecha: $f^1, f^2, f^3, f^4$ y $f^5$ . . . . .	117
5.8.	Esquema para explicar el efecto que tiene LGN en los campos receptivos de V1. . . . .	118
5.9.	Izquierda. Histograma de los errores del ajuste de los campos receptivos según el modelo sigmoide con $\lambda = 0.5$ . Derecha. Histograma de los errores del ajuste de los campos receptivos según el modelo sigmoide con $\lambda = -0.5$ . . . . .	119

5.10. Izquierda. Histograma de los errores del ajuste de los campos receptivos según el modelo sigmoide con $\lambda = 0$ . Derecha. Histograma de los errores de los ajustes de campos receptivos de células de V1 en el macaco (figura extraída de [39]). . . . .	120
5.11. Distribución del ajuste de los campos receptivos en el plano $(n_x, n_y)$ según el modelo sigmoide con $\lambda = 0$ . . . . .	121
5.12. Izquierda. Histograma de las fases del ajuste de los campos receptivos según el modelo sigmoide con $\lambda = 0$ . Derecha. Histograma de las fases de los campos receptivos de células de V1 en el macaco (figura extraída de [39]). . . . .	122
5.13. Comparación de la estructura de los campos receptivos entre datos experimentales [39] y dos modelos teóricos. . . . .	123
A.1. Representación esquemática de la diferencia entre $c$ y $E_0$ . Sólo en el entorno $\Omega$ ambas funciones coinciden. . . . .	130
B.1. Diferentes representaciones de la ecuación B.2.1 con $A = 1/\pi$ , $\Omega = 1$ , $\sigma = 1/2$ , $\varphi = 0$ y $f_y = 0$ como función de $f_x$ para distintos valores de $f_0$ . . . . .	135
B.2. Frecuencia preferida (aquella que maximiza la ecuación B.2.1) como función $f_0$ con $\varphi = 0$ . . . . .	136



# Índice de cuadros

2.1.	Eficiencias $\epsilon_n[c]$ para representaciones con $n = 1, 2$ y $3$ voces para las imágenes <i>imk03236.imc</i> y <i>Lena</i> . . . . .	49
2.2.	Error medio de la condición de ortogonalidad para la wavelet óptima para las componentes $r = 0, 1$ y $2$ . . . . .	49
3.1.	Tabla doble: densidad (arriba), supercaracterística visual a la escala de detección (centro) y a escala doble (abajo), ordenadas por sus abundancias relativas dentro del colectivo. . . . .	83
4.1.	Distribución de los signos de los $\alpha_{0\vec{0}}$ (expresados en porcentaje) de las proyecciones en seis bases de wavelets y para un subcolectivo de 1000 imágenes de van Hateren. Las etiquetas $H$ , $V$ y $D$ corresponden, respectivamente, a las wavelets Horizontal, Vertical y Diagonal, respectivamente	88
4.2.	Promedios de los $\eta$ , de sus signos ( $\beta$ ), de sus módulos ( $ \eta_j $ ) y de los $\lambda_j$ para distintas escalas para el colectivo de las 4167 imágenes. . . . .	94